NASA Conference Publication 2504

Mental-State Estimation 1987

(NASA-CP-2504) MENTAL-STATE ESTIMATION, 1987 (NASA) 393 p CSCL [illegible]
N88-23370
--THRU--
N88-23397
Unclas
[illegible]/53 0146798

Proceedings of a workshop held in
Williamsburg, Virginia
June 3-4, 1987

NASA

NASA Conference Publication 2504

Mental-State Estimation 1987

Compiled by
J. Raymond Comstock, Jr.
Langley Research Center
Hampton, Virginia

Proceedings of a workshop sponsored by the
National Aeronautics and Space Administration,
Washington, D.C., and Old Dominion University,
Norfolk, Virginia, and held in
Williamsburg, Virginia
June 3–4, 1987

NASA
National Aeronautics
and Space Administration

Scientific and Technical
Information Division

1988

FOREWORD

The Mental-State Estimation Workshop - 1987 was held June 3 to 4, 1987, in Williamsburg, Virginia. The workshop was sponsored by the Human Engineering Methods Group, Crew/Vehicle Interface Research Branch, Flight Management Division, NASA Langley Research Center, and the Center for Ergonomics Research and Training, Old Dominion University, Norfolk, Virginia.

A total of 78 persons attended the workshop; 29 of these individuals gave presentations at the workshop. Presenters and attendees represented the government, corporations, and universities.

One purpose of the workshop was to examine the status of the idea that cognitive and emotional processes, or mental states, are reflected in discriminable patterns of physiological response. The intent of the workshop was to explore further the potential of a technology based on this concept and the projected flight management applications for such a technology. These technology applications would contribute to the goal of facilitating crew performance of flight management tasks.

Each person presenting a paper at the workshop was asked to consider the following questions and explain how they might be relevant to their work or work area.

- In what ways is the concept of mental states useful in your work and in aerospace human factors work in general?
- What are the mental-state constructs with which you work, how are they operationally defined, and what are your working hypotheses regarding the relationships between them?
- In what ways do you recommend that the issue of individual differences or trait differences be taken into account?
- What dimensions of similarity between laboratory analog tasks and operational environments to which results must transfer (such as the aerospace flight deck) do you consider important?
- What are the strengths and weaknesses of the analytical methods that you use and what would be the characteristics or capabilities of the ideal method for your purposes?
- What are your views concerning the characteristics and capabilities of an ideal technology for evaluating the "man-machine interface?"

J. Raymond Comstock, Jr.

PRECEDING PAGE BLANK NOT FILMED

ACKNOWLEDGMENTS

The Mental-State Estimation Workshop - 1987 would not have been possible without the contributions of each person in the Human Engineering Methods Group, Crew/Vehicle Interface Research Branch, NASA Langley Research Center. The following people are recognized for participation in the planning and conduct of the workshop: Randall L. Harris, Sr., Alan T. Pope, Mark Nataupsky, Daniel W. Burdette, Gregory A. Bonadies, and Mary L. McManus.

Appreciation for support and guidance is extended to Jack J. Hatfield, Branch Head, Crew/Vehicle Interface Research Branch, Samuel A. Morello, Assistant Division Chief, and John F. Garren, Division Chief, Flight Management Division.

CONTENTS

SCHEDULE OF EVENTS:

MENTAL-STATE ESTIMATION WORKSHOP - 1987

Williamsburg, Virginia

June 3-4, 1987

(Updated Titles for Proceedings)

Introductory Remarks: Dr. Alan T. Pope (NASA Langley Research Center)

Keynote address: Dr. Gary E. Schwartz (Yale University)

SESSION I: Physiological measures of mental state in operational settings: Current approaches and future challenges

Chair: Richard L. Horst (ARD Corporation)

Horst, R. L. An overview of current approaches and future challenges in physiological monitoring S1

Samaras, G. Towards a mathematical formalism of performance, task difficulty, and activation S2

Porges, S. Vagal tone as an index of mental state S3

Banta, G. Challenges of physiological monitoring in a Navy operational setting S4

Yates, R. U.S. Air Force techniques for measuring mental state omit

Aldrich, T. B. Predicting operator workload during system design S5

SESSION II: Stress and Stress effects

Chair: R. P. Bateman (Boeing)

Alkov, R. A. Chronic stress as a factor in aircraft mishaps S6

Bateman, R. P. Acute Stress S7

SESSION VI: State of the art methods, technologies and applications of neurophysiological predictors of quality of performance

Chair: Alan Gevins (EEG Systems Laboratory)

Gevins, A. Neurophysiological predictors of quality of performance S21

SESSION VII: Prediction and Biocybernetics

Chairs: Robert O'Donnell (NTI, Inc.), & Sam Shifflet (Brooks AFB)

Stern, J. Physiological measures and mental state assessment S22

Shingledecker, C. A correlational approach to predicting operator status S23

Gilliland, K. Brainstem response and state-trait variables S24

Brenner, M. Voice stress analysis

Strome, D. Development of a C^3 generic workstation: systems overview S26

Eddy, D. C^3 generic workstation: Performance metrics and applications S27

OMIT TO
P. 25

KEYNOTE ADDRESS

Gary E. Schwartz of Yale University delivered the keynote address at the Mental-State Estimation Workshop - 1987. Because a transcript of the keynote address would lose meaning without the many slides and viewgraphs presented by Dr. Schwartz, a reprint of his chapter "Emotion and Psychophysiological Organization: A Systems Approach," appears here. The chapter appeared in Coles, Donchin, and Porges, Psychophysiology, 1986, The Guilford Press. The chapter is reprinted with the permission of Dr. Schwartz and The Guilford Press.

Chapter Seventeen

Emotion and Psychophysiological Organization: A Systems Approach

Gary E. Schwartz*

INTRODUCTION AND OVERVIEW

The topic of emotion is one of the most fundamental and confusing areas in psychophysiology. It is a common belief among psychophysiologists (as well as laypersons) that bodily processes are related to emotional experiences and expressions, and that this relationship is fundamental to biological, psychological, and social well-being. However, the nature of the relationship between bodily processes and emotional experience and expression is not well understood. The literature suffers from conceptual and methodological problems that inadvertently encourage continued confusion rather than clarity. The purpose of this chapter is not only to provide a selective review of the recent research on the psychophysiology of emotion, but also to propose a conceptual framework having direct methodological implications that promises to bring light and clarity to this fundamental area.

In Greenfield and Sternbach's (1972) *Handbook of Psychophysiology*, the chapter on emotion emphasized that "in human subjects, emotional behavior includes responses in three expressive systems: verbal, gross motor, and physiology (autonomic, cortical, and neuromuscular)" (Lang, Rice, & Sternbach, 1972, p. 624). Lang *et al.* further emphasized that "the responses of no single system seem to define or encompass an 'emotion' completely." In the 13 years that have passed since this important chapter was written, some progress has been made in describing the empirical relationship among these "three expressive systems" (e.g., see Lang, Miller, & Levin, 1983). However, little progress has been made in clarifying the conceptual relationship among these three expressive systems (Schwartz, 1978, 1982).

I propose that general systems theory (deRosnay, 1979; von Bertalanffy, 1968) provides a framework for understanding the relationship between the concept of emotion and the various measurable components presumed to reflect the presence of emotion. As becomes clear as the chapter unfolds, the concept of emotion is an inferred concept, not unlike inferred concepts from modern physics. The concept of emotion is evoked to explain why it is that subjective experience, overt behavior and physiology are at times *organized* and *coordinated* to achieve particular organism–environmental interactions. In fact, it is reasonable to propose that the concept of emotion, appropriately defined, can be a fundamental organizing principle in psychophysiology. Simply stated, emo-

* Gary E. Schwartz. Department of Psychology, Yale University, New Haven, Connecticut.

tion may be the process whereby the three expressive systems (or more appropriately stated, subsystems) are organized in order to achieve specific biopsychosocial goals (Schwartz, 1984).

The central thesis of this chapter is that emotion reflects a fundamental mechanism whereby biopsychosocial processes are organized to achieve specific adaptive goals. After considering various theories of emotion from a systems perspective, I review recent research on patterns of subjective experience, patterns of skeletal muscle activity, patterns of autonomic activity, and patterns of central nervous system activity. The relationship of emotion to personality and to disease is also considered. Finally, implications of viewing emotion as an organizing concept for research methods in psychophysiology and emotion are discussed.

EMOTION FROM A SYSTEMS PERSPECTIVE: THE IMPORTANCE OF ORGANIZED PATTERNING AND EMERGENT PROPERTIES

A fundamental tenet of systems theory is that a system is a "whole" composed of a set of "parts" (i.e., subsystems). The parts interact in novel ways to produce unique properties or "behaviors" of the system as a whole. Therefore, the behavior of a system is said to "emerge" out of the interaction of its parts. The concept of the behavior of a whole system being qualitatively different from the simple sum of the behavior of its parts, yet being dependent upon the interaction of its parts for its unique properties as a whole, is very general. This concept can be applied to any system, be it living or nonliving, be it at a micro level (such as the atom) or at a macro level (such as the social group) (von Bertalanffy, 1968).

Although the general concept of emergent property is by no means fully understood or free from controversy (Phillips, 1976), it is nonetheless considered by most philosophers of science to be fundamentally true. Emergent phenomena are found at all levels in nature, from mathematics and physics, through chemistry and biochemistry, to biology and psychology, sociology, political science, and beyond (e.g., ecology and astronomy).

One difficulty in thinking across levels of complexity (and, therefore, across disciplines) is that one discipline's "system" often turns out to be another discipline's "part." For example, for the physiologist the "system" is "physiology," which itself is composed of parts (organs are composed of cells), whereas for the psychologist the "physiology" becomes the parts that comprise a person or lower animal (organisms are composed of organ systems). We can apply this issue to the relationship between physiology and emotion. From a systems point of view, emotion at the organism level emerges out of the interaction of biological parts at the physiological level. From this perspective, the "behavior" of the physiology is not a "correlate" of the "emotion," regardless of where the physiology is measured (peripherally or centrally). Rather, the physiology should be viewed and described as being a "component" of the "emotion" in the same way that a cell is considered to be a component (rather than a correlate) of an organ.

Thinking in systems terms can be confusing, because words such as "behavior" and "level" must be carefully redefined. The systems theorist would argue that it is as reasonable to speak of the behavior of a nerve, or the behavior of a muscle, as it is to speak of the behavior of a person, or the behavior of a group. "Behavior" is an abstract concept that applies to any level in any system. Consequently, when a person "behaves" at a psychological level, he or she is also "behaving" at a physiological level (and every level below this). In systems terms, it is imprecise to say that tensing the muscles in one's arm is a "correlate" of overt movement behavior; rather, it is a "component" of the overt behavior, and furthermore, it is itself a behaving process! The reason why *Behavioral Science*, the journal of the Society for General Systems Research, publishes selected articles in physics and physiology as well as in psychology and sociology is that it adopts the concept of behavior as being very general—a concept that can be applied to any system at any level. To use the term "behavior," then, requires that it be carefully qualified regarding the level of analysis (see Table 17-1). How this applies to the psychophysiology of emotion is explained shortly.

There is a tricky problem in defining levels, because different levels occur *within* disciplines as well as *across* disciplines. For example, in psychology, one can speak of complex cognitive processes as being composed of underlying component cognitive processes (Sternberg, 1977) in the same way that in physiology one can speak of complex cardiovascular processes as being composed of underlying component physiological processes (Miller, 1978; Schwartz, 1983). Note that specifying such sublevels *within* a given discipline does not eliminate the concept of unique properties (behaviors) emerging out of components interacting with one another. Rather, the need to specify levels within a given discipline (as well as across disciplines) requires that we think more clearly about what is really a component of what.

The implications of levels and emergent properties for the psychophysiology of emotion are important. Although it was an essential first step to describe emotion as consisting of three basic components (subjective experience, overt behavior, and physiological activity; reviewed in Lang *et al.*, 1972), it is a mistake to think that these three categories operate at the same level of analysis, and therefore to treat the three cate-

Table 17-1
Levels of Complexity in systems and Associated Academic Disciplines

Level and complexity of the system	Academic discipline associated with the level of the system
Beyond earth	Astronomy
Supranational	Ecology
National	Government, political science, economics
Organizations	Organizational science
Groups	Sociology
Organism	Psychology, ethology, zoology
Organs	Organ physiology, (e.g., neurology, cardiology)
Cells	Cellular biology
Biochemicals	Biochemistry
Chemicals	Chemistry, physical chemistry
Atoms	Physics
Subatomic particles	Subatomic physics
Abstract systems	Mathematics, philosophy

Note. According to systems theory, in order to understand the behavior of an open system at any one level, it is essential to have some training in the academic disciplines below that level, plus have some training in the relevant discipline at the next highest level as well.

From "A Systems Analysis of Psychobiology and Behavior Therapy: Implications for Behavior Medicine" by G. E. Schwartz, *Psychotherapy and Psychosomatics*, 1981, 36, 159-184. Copyright 1981 by *Psychotherapy and Psychosomatics*. Reprinted by permission.

gories as if they were relatively independent parts. First of all, from a systems point of view, subjective experience and "overt" behavior (note that the word "behavior" is qualified here as required by systems theory) are both categories of behavior at the organism level, each of which is comprised of patterns of physiological processes. Physiological processes are therefore not independent of these two categories. On the contrary, physiological processes are the building blocks of both of these processes, and must therefore be conceptualized and researched from this perspective.

Moreover, from a systems perspective, subjective experience and overt behavior are themselves not at the same level. Whereas subjective experience is completely personal (private at the organism level), overt behavior is fundamentally social (i.e., it allows the organism to communicate and interact with the environment of which the organism is a part). Hence, a systems approach leads to the suggestion that emotion is truly a biopsychosocial process (e.g., Engel, 1977; Leigh & Reiser, 1980; Schwartz, 1983) or a social psychobiological process (e.g., Cacioppo & Petty, 1983), depending upon whether one views the process from the micro to the macro levels (biopsychosocial) or from the macro to the micro levels (social psychobiology). In both cases, subjective experience and cognitive processing become the "middle" processes between one's biology (at the micro level) and one's social interactions (at the macro level).

Note that from a systems perspective, patterns of processes can occur at each level (biological, psychological, and social). Interactions at each level should lead to emergent properties at the next level (e.g., physiological patterns should contribute to unique subjective experiences, and subjective patterns should contribute to unique social interactions). From a systems perspective, not only does psychology emerge out of physiology, and social behavior emerge out of psychology, but physiology, psychology and social behavior represent different levels on analysis *of the same, ultimate, whole system.* It should be clear that according to systems theory, analyzing the physiological parts in relative isolation, or analyzing the subjective or social parts in relative isolation, will not lead to a complete understanding of emotion as a whole process. Emotion takes on its unique holistic properties as a result of complex *interactions* and *organizations* of its component processes *at each level.* This is why an analysis of emotional processes from a systems point of view requires that the investigator measure patterns of variables across levels and search for unique *interactions* or *emergents* between and among the variables across levels.

One point needs to be underscored here before we move to empirical findings. A fundamental question that needs to be addressed is this: "How is it that biological, psychological, and social processes are ever organized?" Where does the apparent order come from? Clearly, there are instances where physiological, subjective, and behavioral responses are relatively dissociated (e.g., Weinberger, Schwartz, & Davidson, 1979). However, implicit in the concept of emotion is the notion that a fundamental organization does exist and has evolutionary, adaptive significance for survival (Darwin, 1872; Plutchik, 1980). If there was no order, no organization of biological, psychological, and social processes, the study of emotion would ultimately be impossible. Moreover, there would be no empirical utility or scientific justification for having a concept of emotion.

Systems theory encourages psychophysiologists to look for organized patterns of processes, within and across levels of nature. The concept of emotion, when viewed from the perspective of systems theory, has the potential to clarify the growing literature indicating that organized patterns of processes can and do occur within and across the biological, psychological, and social levels. Moreover, the concept of emotion, when viewed from the perspective of systems theory, has the potential to clarify and stimulate new methods

for measuring emotion at physiological, subjective, and social levels.

It is difficult to discuss research on human emotion without beginning in the middle—that is, at the psychological level. Researchers and laypersons alike often begin with their own subjective experience, and then look for relationships among their subjective experience, biology, and overt behavior. Although a systems approach to emotion encourages us to take a comprehensive, biopsychosocial approach to emotion, this chapter focuses primarily on the relationship between the psychological and biological levels, and, within the biological level, primarily on the physiological (organ) level (excluding the cellular and biochemical—e.g., neurohumoral and neuroendocrine—levels). Since this volume is concerned with human psychophysiology, the present chapter reviews representative empirical findings in psychology first and then considers relationships between psychology and physiology.

Most modern theorists of emotion believe that different emotions reflect different organizations or patterns of processes at psychological and physiological levels of analysis (e.g., Izard, 1977; Lang *et al.*, 1972; Leventhal, 1980; Plutchik, 1980; Tompkins, 1980), though some modern theorists focus primarily on the psychological level (e.g., Zajonc, 1980). It is important to recognize that the concept of patterning of psychological and biological processes is explicitly made in many classic theories of emotion (Darwin, 1872; James, 1884) and is implicitly made in more recent social psychobiological theories of emotions (e.g., Schachter & Singer, 1962).

Schachter and Singer (1962) proposed that emotional experience and emotional social behavior reflect an interaction of the nature of the social situation (e.g., an experimenter makes jokes), the way in which the social situation is interpreted by the subject (e.g., the subject perceives the jokes as humorous), and the subject's level of general physiological arousal (e.g., the subject is aroused by an injection of epinephrine, and, moreover, the subject is not told that the specific physiological side effects of injection are related to the injection itself). Though Schachter and Singer (1962) did not emphasize patterning of processes *within* each of the levels (e.g., they assumed that physiological patterning played little, if any, role in emotional experience and expression), they did emphasize patterning of processes *across* levels. The concept of patterning, be it within and/or between levels of processes, is implicit if not explicit in all theories of emotion.

Before we can meaningfully consider patterning of physiological processes, it is essential to examine some of the recent data on patterning of subjective experience. As becomes clear below, it is possible to measure reliable patterns of subjective experience when the appropriate theoretical and associated methodological considerations are adopted.

PATTERNING OF SUBJECTIVE EXPERIENCE AND EMOTION

Few researchers have systematically examined subjective experience closely to uncover possible distinct patterns *within* the experience as a function of different emotions. The pioneering research of Izard (1972) is very important in this regard. Not only did Izard assess simultaneously multiple emotional experiences to different affective situations using the Differential Emotions Scale (DES), but he proposed that a subset of emotions, such as anxiety and depression, are themselves composed of different combinations of underlying fundamental emotions. Izard has essentially proposed that *within* the psychological level, it is possible to uncover levels of emotional experience, where higher levels of subjective experience presumably represent emergent combinations or patterns of lower levels of subjective experience. In systems terms, what Izard has proposed is that anxiety and depression are each unique emotional states that emerge out of the interaction of patterns of fundamental emotions. At least six different fundamental emotions (happiness, sadness, anger, fear, surprise, and disgust) have been found to exist cross-culturally and to be linked to specific facial expressions of emotion (Ekman, Friesen, & Ellsworth, 1972; Izard, 1971).

As part of a research program at Yale University examining affective imagery and the self-regulation of emotion, we decided to determine whether it was possible to discover standardized situations that college students could imagine that would evoke consistent patterns of subjective experience. In the process of conducting the research to address this basic methodological question, we attempted to replicate and extend Izard's (1972) research on the relationship between the hypothesized higher-order emergent emotions of anxiety and depression, and patterns of fundamental emotions, using an abreviated DES scale (Schwartz & Weinberger, 1980).

Initially, 55 subjects filled out a questionnaire asking them to "give a one-sentence statement or a single phrase about a situation that either happened in the past, or could happen in the future, that would make you feel one of the following: happy, sad, angry, fearful, anxious, depressed." Subjects were further told to note that "for each emotion, three separate situations are requested that reflect three different intensities of emotion: strong, moderate, and weak." Therefore, each subject was required to generate 18 emotional situations.

From this sample, 20 of the questionnaires (from 10 males and 10 females), which were complete and did not contain highly idiosyncratic or redundant answers, were chosen to be validated in a second questionnaire. The items were edited into complete sentences. Then, the 18 items from each of the 20 questionnaires were combined to create a pool of 360

statements. These statements were randomly sorted into four forms of 90 items each, with each emotional category and intensity represented by 5 items per form. A total of 216 subjects filled out one of the forms of the questionnaire, using the following instructions:

> For each of the following statements, *imagine* that they are happening to you, and rate *how you would feel.* Note that each statement has *six* emotions to be *separately* rated—happiness, sadness, anger, fear, depression, and anxiety. Since it is not uncommon for people to experience more than one emotion in a given situation, you should rate each statement on all six emotions. Use the numbers 1–5 for your ratings, with 1 meaning very little, 3 meaning moderate, and 5 meaning very strong. Numbers 2 and 4 should also be used to reflect intermediate categories between very little and moderate, and moderate and very strong, respectively.

As can be seen in Figure 17-1, the average ratings (across the three intensities of items) yielded highly distinct patterns of subjective experience for the four fundamental emotions (Part A of the graph) and similar yet distinct patterns of response comparing sadness with depression and anxiety with fear (Part B of the graph). The richness of these data should not go unnoticed. For example, it can be seen that anger situations elicited more feelings of depression than did either fear or anxiety situations. Also, note that fear situations elicited more feelings of anxiety than anxiety situations elicited feelings of fear.

Figure 17-2 presents the mean ratings for happiness situations subdivided by intensity (high, moderate, low). This figure illustrates not only that the situations reliably elicited primarily feelings of happiness that vary with intensity, but that the *higher* the happiness, the *lower* the sadness, depression, and anger, but *not* the fear and anxiety. In fact, high happiness was accompanied by moderately high feelings of anxiety!

These results can be compared with those obtained for the mean ratings for anxiety situations subdivided by intensity. Figure 17-3 shows not only that these situations reliably evoked primarily feelings of anxiety, but that the *higher* the anxiety, the *higher* the fear, depression, and sadness, while the relationship with happiness and anger is less clear. Happiness items and anxiety items clearly differed in the *patterns* of emotions that they elicited.

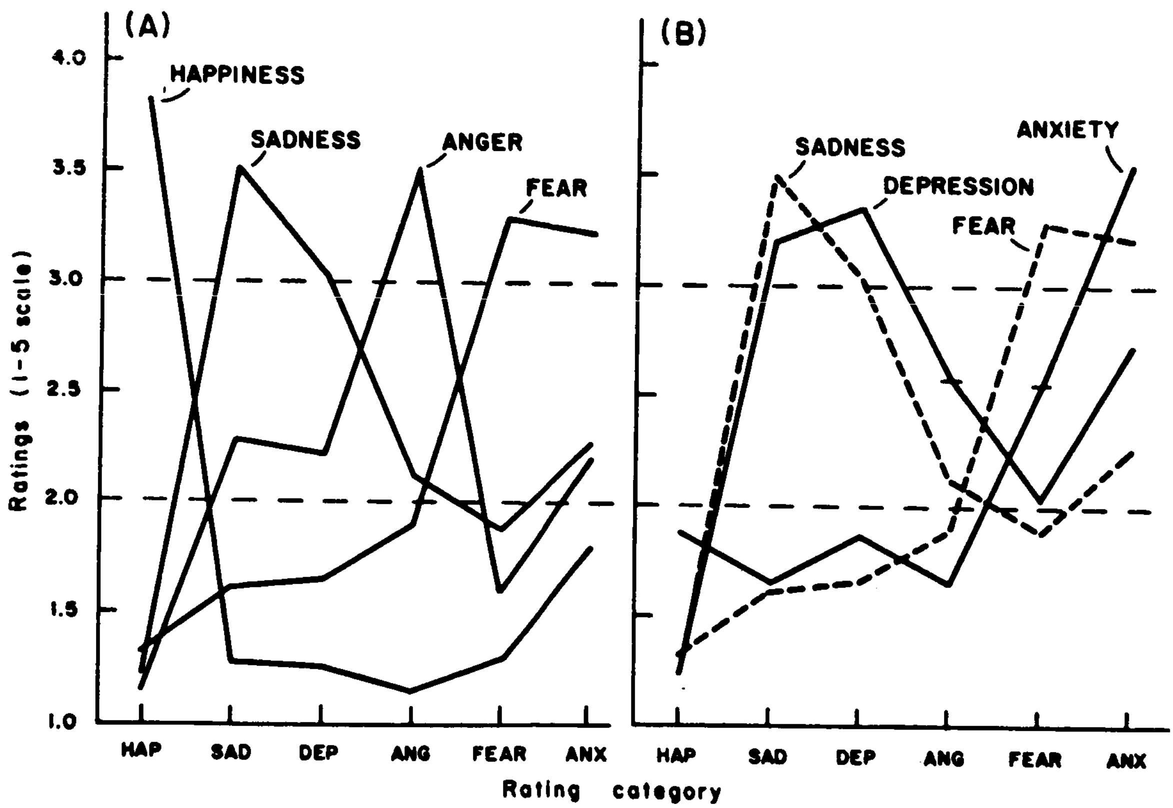


Figure 17-1. Mean ratings of happiness (HAP), sadness (SAD), depression (DEP), anger (ANG), fear (FEAR), and anxiety (ANX) separately for happiness, sadness, anger, and fear situations (1A) and for depression and anxiety situations (1B) (with sadness and fear situations redrawn for comparsion). (From Schwartz and Weinberger, 1980.)

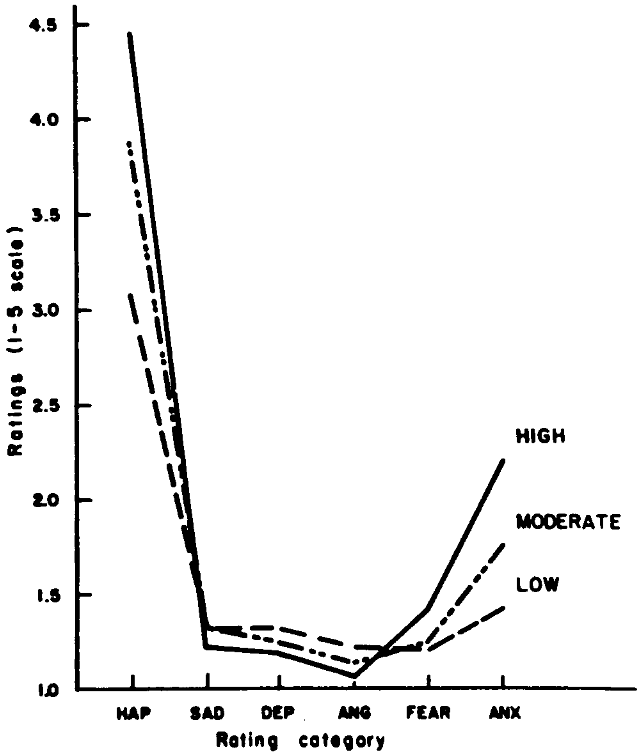


Figure 17-2. Mean ratings for happiness situations subdivided by intensity (high, moderate, low). (From Schwartz and Weinberger, 1980.)

Apparently, at least for a sample of college students, specific classes of affective situations evoked specific patterns of subjective experience. This does not mean that all subjects gave (or give) identical responses to the average items in a given category or to specific items. On the contrary, individual differences in response to standardized situations are of fundamental importance for basic research and clinical applications, and should be assessed carefully. The question of individual differences is discussed later in the context of relating patterns of subjective experience to patterns of physiological activity. The important point to remember here is that distinct patterns of subjective experience to specific classes of affective situations can be assessed and reveal rich complexity and organization in the psychological structure of emotional experience.

The most surprising and informative aspects of findings are uncovered when patterns of responses to individual items are examined (from Schwartz, 1982). It turns out that specific items evoke particular blends or patterns of subjective experience. As shown in Table 17-2, when college students imagined that "Your dog dies," high ratings occurred primarily in sadness and depression (high ratings are in italics in the table), whereas when they imagined that "Your girlfriend/boyfriend leaves you for another," high ratings now occurred in anger and anxiety as well as in sadness and depression. Whereas the former item might be globally labeled as a "sadness" or "depression" item, the latter item (having the same sadness rating) might be globally labeled either as a "sadness" or "depression" item, or as an "anger" item, or as an "anxiety" item (if one views the data in either–or categories). Note that in response to the question "You realize that your goals are impossible to reach," college students rated this situation highly in all of the five negative emotions assessed in the study. Clearly, this particular situation is a complex, highly negative, patterned emotional state. Its relevance to social/political problems facing children, adolescents, and adults in modern society (with the increased societal recognition that fundamental limits do exist and that one's life style must be limited accordingly if society is to survive) should be self-evident. We should not be surprised that the general mood in modern societies today is conflicted, since the pressing social problems probably elicit complex, yet organized blends of fundamental emotions.

These particular patterns of subjective experience are not necessarily unique to Yale University students. It is conceivable that the relative differences in patterns observed among the different situations may

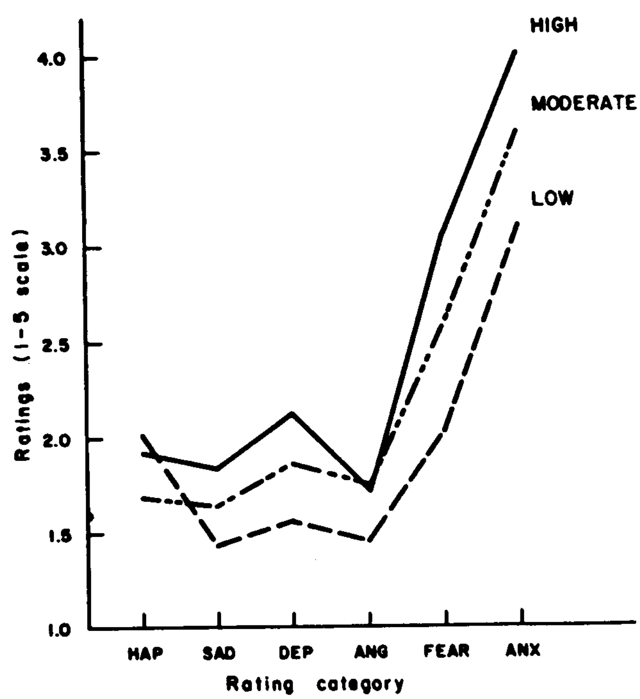


Figure 17-3. Mean ratings for anxiety situations subdivided by intensity (high, moderate, low). (From Schwartz and Weinberger, 1980.)

Table 17-2
Ratings on an Abbreviated DES for Yale University Students

Item	Happiness	Sadness	Anger	Fear	Anxiety	Depression
Your dog dies	1.09	<u>4.08</u>	2.08	1.38	1.93	<u>3.34</u>
Your girlfriend/boyfriend leaves you for another	1.13	<u>4.13</u>	<u>3.41</u>	2.11	<u>2.72</u>	<u>4.09</u>
You realize that your goals are impossible to reach	1.15	<u>3.64</u>	<u>3.00</u>	<u>2.48</u>	<u>3.08</u>	<u>3.67</u>

Note. Underlined entries indicate high ratings.

apply to various age groups and may even apply to various modern cultures. Table 17-3 presents the average self-reports of 23 Italian physicians and psychologists to the same questions. The items were presented in Italian by a translator at a scientific meeting in Rome, Italy, in the spring of 1983; the data reflect the percentage of subjects giving a 4 or 5 rating for a given emotion to a given question. The summary data were collected with the aid of the translator at the meeting itself, prior to my presenting the theory and findings obtained in the United States (see Schwartz & Weinberger, 1980). Depsite the differences in age, academic background, culture, and mode of administration, the relative pattern of differences among the three items is preserved. It is conceivable not only that the meaning of these particular situations was similar for the two samples of subjects, but that the two samples of subjects interpreted the meanings of the emotional words in a similar fashion. One could hypothesize that American college students and Italian professionals would show comparable patterns of physiological responses differentiating among the various emotional situations. However, cross-cultural research comparing subjective and physiological patterns of response to different fundamental emotions and blends of emotions has yet to be reported in the literature.

The value in assessing blends or patterns of subjective experience in the study of emotion should be emphasized for its methodological as well as its theoretical implications. Consider the following ratings on two items that might both be globally described as "high-happiness" items for students at Yale: "You are accepted at Yale" and "You have just graduated from Yale." As shown in Table 17-4, the first item evoked high ratings not only in happiness but also in anxiety. Hence, from an either–or perspective, one could conclude that this item was an "anxiety" item rather than a "happiness" item; clearly, the situation evoked both anxiety and happiness—a blended/patterned emotion. Note that in contrast, the second item evoked moderate to high ratings in sadness, fear, and depression, as well as in happiness and anxiety. Clearly, both situations evoked high "happiness" (and high anxiety) in the average Yale student, but the patterning of emotions in the second item is even more complex than the relatively "pure" happiness of the first item.

This difference in patterns of emotional experience between being admitted to a university versus graduating from a university is apparently not unique to Yale students. As illustrated in Table 17-5, similar differences in relative patterns of emotional experiences were reported by Italian professionals when they compared beginning versus completing their grad-

Table 17-3
Ratings on an Abbreviated DES for Italian Professionals

Item	Happiness	Sadness	Anger	Fear	Anxiety	Depression
Your dog dies	0	<u>52</u>	13	0	17	<u>26</u>
Your girlfriend/boyfriend leaves you for another	0	<u>70</u>	<u>61</u>	13	<u>43</u>	<u>70</u>
You realize that your goals are impossible to reach	0	<u>61</u>	<u>61</u>	<u>30</u>	<u>70</u>	<u>74</u>

Note. Each entry represents the *percentage* of subjects giving a 4 or 5 rating for a given emotion to a given question. Underlined entries indicate high percentages.

Table 17-4
Ratings on an Abbreviated DES for Yale University Students

Item	Happiness	Sadness	Anger	Fear	Anxiety	Depression
You are accepted at Yale	4.18	1.14	1.04	1.96	3.04	1.09
You have just graduated from Yale	4.09	2.74	1.38	2.57	3.40	2.36

Note. Underlined entries indicate high ratings.

uate training (implying that the findings generalize across undergraduate and graduate schooling, as well as across age, academic background, and culture).

It is fascinating how what at first glance might appear to be minor differences in wording can dramatically change the pattern of subjective experience elicited by an item. As can be seen in Table 17-6, in response to the item "You feel loved," a pure emotion of happiness was generated in Yale college students, whereas in response to the item "You meet someone with whom you fall in love," the more complex pattern of happiness and anxiety was elicited (a more "stressful" situation). Interestingly, as can be seen in Table 17-7, similar relative patterns of subjective experience were reported for the Italian professionals (with the Italians giving less of a pure happiness rating for the second item compared to the first). The shift in wording from feeling loved to meeting someone with whom you fall in love seems to have had similar significance for the two samples of subjects.

There are numerous conclusions that can be drawn from data such as these. It appears that different situations can evoke different combinations of emotions as assessed through self-report. Therefore, if only a single emotion is assessed (a still reasonably common research practice), this will lead to an incomplete if not an erroneous description of the emotional state of the person. The fact that combinations of emotions can be elicited reliably by affective imagery and can be assessed reliably using a simple self-report DES procedure indicates that future research should adopt a pattern approach to assessing and interpreting the subjective dimensions of emotion. Statistical techniques for assessing patterns of physiological activity using multivariate pattern recognition and classification procedures (Fridlund, Schwartz, & Fowler, 1984; Schwartz, Weinberger, & Singer, 1981), to be discussed later in the sections on patterning of skeletal and autonomic muscle activity, can be similarly applied to assessing patterns of subjective experience. From a systems point of view, the conceptual approach to assessing patterns of responses should be sufficiently general to apply to all levels and disciplines.

Are different patterns of subjective experience associated with different patterns of physiological responses? Are the weak and often inconsistent findings in the psychophysiological literature linking subjective experience to patterns of physiological activity due, at least in part, to the fact that patterns of subjective experience have not been assessed? If patterns of subjective experience are assessed, will we find that certain situations are better than others in eliciting relatively pure emotions? Do emotions actually occur simultaneously in patterns, or do fundamental emotions shift from one to another, whereas the subjective impression is that they occur concurrently? These questions and many others are stimulated when one

Table 17-5
Ratings on an Abbreviated DES for Italian Professionals

Item	Happiness	Sadness	Anger	Fear	Anxiety	Depression
You are accepted at school	83	0	4	13	26	4
You have just graduated from school	61	26	0	39	70	22

Note. Each entry represents the *percentage* of subjects giving a 4 or 5 rating for a given emotion to a given question. Underlined entries indicate high percentages.

Table 17-6
Ratings on an Abbreviated DES for Yale University Students

Item	Happiness	Sadness	Anger	Fear	Anxiety	Depression
You feel loved	4.78	1.28	1.13	1.19	1.57	1.19
You meet someone with whom you fall in love	4.58	1.20	1.04	2.00	3.06	1.33

Note. Underlined entries indicate high ratings.

begins to adopt a systems perspective and applies the perspective to the study of patterns of biopsychosocial responses in emotion.

SKELETAL MUSCLE PATTERNING AND EMOTION

If any single physiological system is designed to express different emotions, it is the skeletal muscle system. The skeletal muscles can be finely regulated by the brain to produce delicate, precise, and highly complex patterns of activity across muscles and over time. The face, with its high ratio of single motor units to muscle mass, and its rich neural innervation, is a muscular system anatomically and neurally capable of reflecting different fundamental emotions and patterns of emotions (Ekman & Friesen, 1978).

Whether one chooses to label facial expression as "psychological" behavior or "physiological" behavior is more a reflection of the orientation of the observer than it is a true psychophysiological distinction (Schwartz, 1978). In systems terms, what we observe overtly as facial expression *is* an indirect indicator of complex patterns of facial muscle activity. This is the basis of the comprehensive, anatomically derived visual rating system for scoring overt facial expression developed by Ekman and Friesen (1978).

Subtle and fast-acting changes in muscle activity can be readily quantified by attaching miniature silver/silver chloride electrodes to the surface of the skin over relative muscle regions (see Figure 17-4). More precise measurements can be made using fine wire needle electrodes inserted through the skin to monitor the activity of single motor units (Basmajian, 1978). Both of these electromyogram (EMG) methods are relatively obtrusive. EMG methods restrict the subject's freedom of movement and often increase the subject's attention to his or her facial behavior. Consequently, EMG recordings may influence the affective processes being measured. A recent chapter by Fridlund and Izard (1983) provides an excellent review of the facial EMG literature and discusses the methodological difficulties involved in conducting such research and interpreting the findings. Despite these complications, important basic and clinical information can be obtained using EMG (so long as the restrictions of the method are kept firmly in mind).

It should be recognized that research on patterns of facial muscle activity (and other skeletal muscle activity) has not been restricted to the study of emotion per se. For example, in the program of research conducted by McGuigan and colleagues (reviewed in McGuigan, 1978) and Cacioppo and Petty (reviewed in Cacioppo & Petty, 1981), different patterns of facial EMG have been associated with different cognitive and social information-processing tasks.

In a series of studies, my colleagues and I have documented the sensitivity of facial EMG patterning in differentiating low to moderate intensity emotional states elicted by affective imagery (Schwartz, Fair,

Table 17-7
Ratings on an Abbreviated DES for Italian Professionals

Item	Happiness	Sadness	Anger	Fear	Anxiety	Depression
You feel loved	100	0	0	22	34	0
You meet someone with whom you fall in love	96	0	0	30	70	0

Note. Each entry represents the *percentage* of subjects giving a 4 or 5 rating for a given emotion to a given question. Underlined entries indicate high percentages.

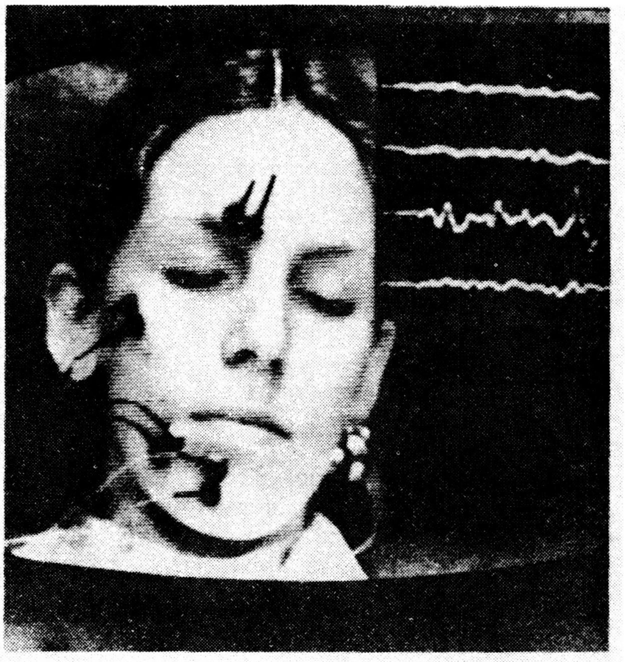


Figure 17-4. Photograph of a videoscreen showing the placement of four pairs of EMG electrodes and, superimposed electronically next to the face, the oscilloscope tracings of the amplified electromyographic activity from the four facial regions. (From Schwartz *et al.*, 1976b.)*

Salt, Mandel, & Klerman, 1976a; Schwartz, Fair, Salt, Mandel, & Klerman, 1976b; Schwartz, Ahern, & Brown, 1979; Schwartz, Brown, & Ahern, 1980). Some of the major results of these studies can be briefly summarized as follows:

1. Different patterns of facial muscle activity accompany the generation of happy, sad, and angry imagery, and these patterns are not typically noticeable in the overt face.
2. Instructions to re-experience or "feel" the specific emotions result in greater EMG changes in relevant muscles than instructions simply to "think" about the situations (see Figure 17-5).
3. Depressed patients show a selective attenuation in the generation of facial EMG patterns accompanying happy imagery, but show a slight accentuation in the facial EMG response to sad imagery (see Figure 17-5). The biggest facial EMG difference between depressed and nondepressed subjects occurs when subjects imagine what they do in a "typical day," with nondepressed subjects generating miniature happy facial EMG pattern and depressed subjects generating a miniature mixed sadness–anger facial EMG pattern.
4. Changes in clinical depression following treatment with active drug medication or placebo are accompanied by relevant changes in facial EMG. Also, higher initial resting levels of facial EMG appear to be predictive of subsequent clinical improvement.
5. Females (compared to males) tend to:

a. Generate facial EMG patterns of greater magnitude (relative to rest) during affective imagery, and report a corresponding stronger subjective experience to the affective imagery.
b. Show greater within-subject correlations between the experience of particular emotions and relevant facial muscles.
c. Show somewhat higher corrugator levels during rest (possibly reflecting more sadness and/or concern) and lower masseter levels during rest (possibly reflecting less anger).
d. Generate larger facial EMG patterns when instructed to voluntarily produce overt expressions reflecting different emotions.

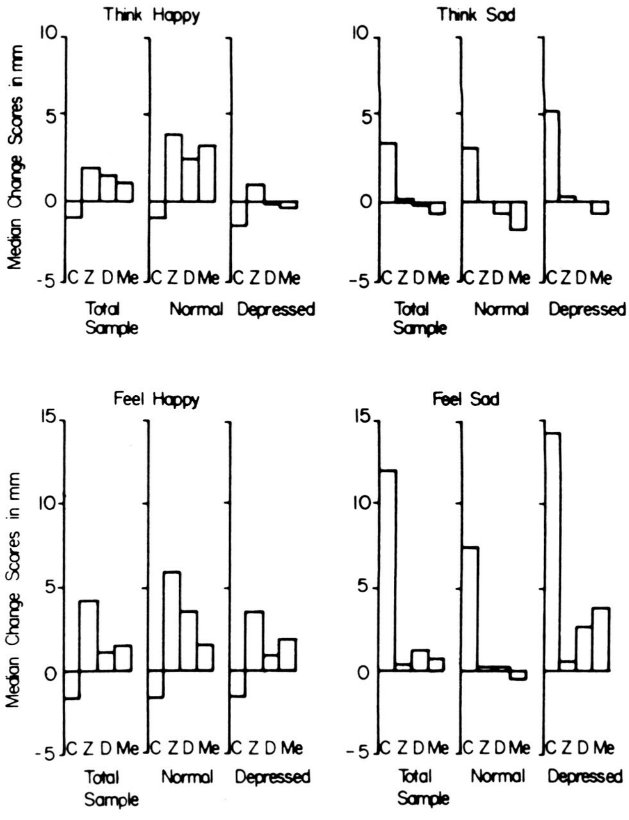


Figure 17-5. Change from baseline for muscle activity from the corrugator (C), zygomatic (Z), depressor annuli oris (D), and mentalis (Me) regions during two affective imagery (happy, sad) and two instructional (think, feel) conditions. Data are displayed separately for the total sample ($N = 24$), the normal subgroup ($N = 12$), and a depressed subgroup ($N = 12$). A 1 mm change score equals 45 μV/30 sec. (From Schwartz *et al.*, 1976b.)

* Original photo not available at time of publication.

Taken together, these data strongly support the hypothesis not only that affective imagery results in reliable self-report of different patterns of subjective experience (reviewed in the previous section), but that these self-reports are *preceded* by the generation of unique patterns of facial muscle activity that vary both in pattern and intensity with the subsequent self-reports. Since the facial EMG situations are usually not visible to an observer, and also are not typically perceived by the subject (whose attention during imagery is largely focused on the images and associated feeling states rather than on his or her face per se), it is reasonable to hypothesize that the self-reports and the facial patterns are reflecting two different aspects of the same, underlying neuropsychological system. This is not to say that self-report and facial activity need always covary or be synonymous. On the contrary, according to systems theory, self-report is an emergent process dependent upon the interaction of multiple processes in addition to facial feedback (both central and peripheral), just as facial behavior is itself an emergent process dependent upon the interaction of multiple neuropsychological processes in addition to the expression of emotion. In systems terms, the concept of a "single" response is an oversimplification, since any "behavior" reflects a composite or pattern of underlying processes. This fundamental point is directly related to the whole–part–emergent concept presented at the beginning of the chapter.

Until discrete patterns of facial EMG are discovered that reflect relatively pure fundamental emotions, it is not possible to address the more complex and intriguing question regarding blends or combinations of different emotions and their relationship to complex patterns of facial EMG. In a recent experiment, we (Polonsky & Schwartz, 1984) attempted to determine whether images designed to evoke a *combination* of happiness and sadness would elicit a *combination* of facial muscle responses previously found to be reliably associated with happiness and sadness. Prior research (e.g., Schwartz *et al.*, 1980) has documented that zygomatic activity increases reliably in happiness, while corrugator activity may simultaneously decrease below resting levels in happiness. This pattern is virtually reversed for sadness: Corrugator activity increases reliably in sadness, while zygomatic activity typically remains at baseline. We (Polonsky & Schwartz, 1984) predicted that items selected to elicit a *combination* of happiness and sadness should be accompanied by relative increases in *both* zygomatic and corrugator activity, though the magnitude of each increase would be less than that found in response to relatively pure emotion items reflecting happiness versus sadness.

In the experiment, a standard pure happy item was "You feel loved"; a standard pure sad item was "Someone close to you dies"; and a standard mixed happy-sad item was "You feel that you are finally separated from your family and are really a tremendous sense of freedom about that, but at the same time you miss the closeness that you had or potential closeness that you could have had." The data indicated that as predicted, the combined happy–sad item generated moderate increases in both zygomatic and corrugator activity, whereas the happy item generated large increases in zygomatic activity unaccompanied by increases in corrugator activity, and the sad item generated large increases in corrugator activity unaccompanied by increases in zygomatic activity. It is important to note that the moderate levels of zygomatic and corrugator activity observed in the combined happy–sad item corresponded to moderate levels of perceived intensity as indicated by self-report for the combined happy–sad item (compared to the happy and sad items, respectively).

These are the first data documenting that discrete emotional blends of affective subjective experience can be associated with discrete blends of physiological activity (i.e., facial EMG). Whether or not more complex blends of affective experience can be mapped onto more complex blends of skeletal muscle activity remains to be determined in future research. Also, the precise timing and synchrony of the blended emotions (e.g., do the emotions actually occur simultaneously in real time as measured by facial EMG, or do they flip back and forth in some cyclic fashion?) remain to be investigated. It is clear that the potential now exists for addressing such questions by taking advantage of advances in the measurement of self-report patterns and EMG patterns.

As discussed elsewhere (Fridlund & Izard, 1983; Schwartz, 1982), the previous research has used relatively simple (i.e., univariate) and therefore conservative statistical procedures for quantifying patterns of responses (be they self-report or physiological). A systems approach to pattern data proposes that more complex, sensitive multivariate statistical analyses should be performed. We (Fridlund *et al.*, 1984) have recently demonstrated how multivariate pattern classification strategies can be applied to facial EMG data (see Figure 17-6). Within this general framework, multiple physiological variables are recorded and digitized by computer (transduced); particular components of each variable are selected for analysis (e.g., means, standard deviations, peaks, time to peaks, etc.); and then a statistical iterative process is performed, whereby specific features are derived that maximally discriminate among sets of variables (feature extraction) and show maximal hit rates on these sets of variables (classification). Using this approach, the reliability of classification hit rates can be used to index the success of the pattern recognition procedure, thereby demonstrating the degree of discriminability of the organization of the input variables.

In the Fridlund *et al.* (1984) experiment, 12 females were administered 48 counterbalanced 20-sec trials of affective imagery, using items preselected for

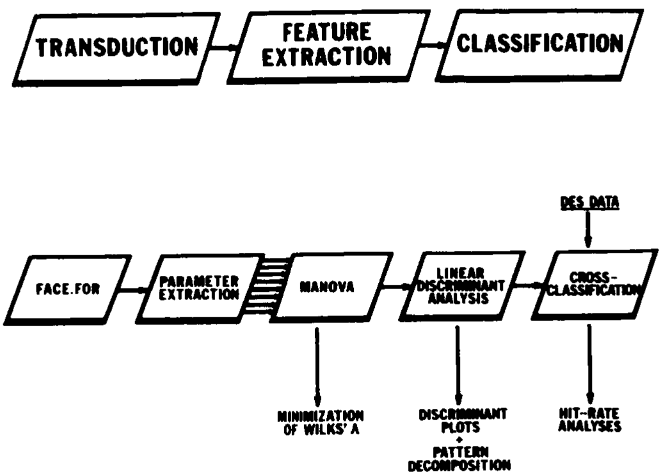


Figure 17-6. Steps used for pattern classification from a systems perspective. (From Fridlund *et al.*, 1984.)

relative purity along the dimensions of happiness, sadness, anger, and fear (from Schwartz & Weinberger, 1980). Facial EMG was recorded from the lateral frontalis, corrugator, orbicularis oculi, and orbicularis oris regions. The findings documented the superiority of statistical strategies that were sensitive to patterns of multiple physiological responses over traditional univariate methods. Moreover, the degree of facial EMB discriminability across emotions within subjects was correlated with the subjects' perceived vividness of their affective imagery.

By using a large number of trials (48) within subjects, it became possible to apply the pattern classification procedures to individual subjects. Figure 17-7 shows a single subject's data comparing anger and fear items for the four separate muscles individually and the composite results of linear discriminant analysis combining the four muscles. This figure illustrates how the multivariate analysis can pull out an anger–fear difference that is not readily apparent in any single muscle.

These multivariate pattern analysis procedures can be applied to patterns of any "input variables," be they self-reports, facial EMG, autonomic responses, electrocortical responses, or patterns across these classes of response systems. The integration of these procedures with research on the psychophysiology of emotion promises to resolve prior confusions and reveal new organized patterns. Unfortunately, to do so requires that we develop new statistical skills and learn new ways of thinking about patterning in systems terms.

One conclusion seems justified from the EMG data available to date: The face is a system that is exquisitively sensitive to underlying affective processes. It therefore provides an excellent window for studying the relationship between subjective experience and physiological activity.

AUTONOMIC PATTERNING AND EMOTION

At one time, it was generally believed that responses innervated by the autonomic nervous system were highly intercorrelated and involuntary, and therefore only capable of reflecting overall levels of arousal and/or alertness (from deep sleep to states of awake excitement). However, it is now well known that the sympathetic and parasympathetic branches of the autonomic nervous system are each capable of very fine regulation of specific peripheral organs. Moreover, this regulation is quite selective and can be brought under voluntary control using such techniques as biofeedback (Schwartz & Beatty, 1977).

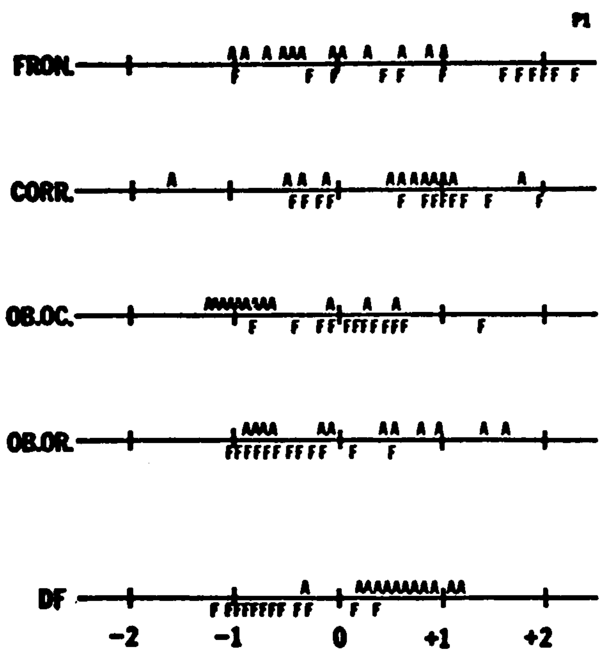


Figure 17-7. Plots of standard EMG scores for 12 anger and 12 fear responses of subject P1 mapped itemwise on each of four muscle regions, and on a linear composite of the four regions derived from linear discriminant analysis. It can be seen that the composite function affords better separation of anger and fear items than any of the individual muscle regions. This figure demonstrates that consideration of variable conformations/patterns provides information which cannot be gleaned from any of the univariate analyses alone. Item codes: A, anger; F, fear; FRON, frontalis; CORR, corrugator; OB.OC, obicularis oculi; OB.OR, obicularis oris; DF, discriminant function. (From Fridlund *et al.*, 1984.)

All physiological responses, to varying degrees, seem to be influenced by both voluntary and involuntary processes. Skeletal muscles are strongly influenced by voluntary processes, but they are also controlled by involuntary reflex patterns elicited by particular stimuli (e.g., in response to localized pain). It now appears that visceral and glandular responses are influenced by voluntary processes more strongly than was previously recognized, though the extent of such control relative to their involuntary reflex patterns is just beginning to be determined.

There has been a paucity of studies examining autonomic patterning accompanying different emotional states. There are many reasons for this relative lack of research. They include methodological problems in recording and analyzing the data, theoretical biases that have discouraged investigators from looking for patterns or accepting evidence of patterning in the data when the patterns emerged serendipitously, and problems at a psychological level in eliciting and assessing the emotional states. However, the few studies that have attempted to address this question have come up with a surprisingly consistent pattern of findings. These studies have focused on the comparison between anger versus fear, two emotions that Ax (1953) claimed were most described as being identical physiological states. The studies prior to 1957 were reviewed by Schachter (1957). A more recent study was reported by Weerts and Roberts (1976).

Drawing on neuropsychological and neuroendocrine findings, Ax proposed that anger involved a *mixed* epinephrine and norepinephrine pattern, while fear involved a relatively pure epinephrine pattern. Schachter added that pain involved a relatively pure norepinephrine pattern, though his pain stimulus (the cold pressor test) may have pulled for this particular response because of its local vasoconstrictive effects.

Unfortunately (though understandably), no single autonomic response is a "pure" reflector of epinephrine- or norepinephrine-like patterns. Most autonomic responses are dually innervated by the sympathetic and parasympathetic branches of the autonomic nervous system, as well as by hormones. For example, an increase in heart rate can by mediated by numerous factors, including (1) an increase in peripheral sympathetic activity, (2) a decrease in peripheral parasympathetic activity, (3) an increase in circulating epinephrine (to list only one possible heart-rate-stimulating hormone), or any combination or pattern of these mechanisms.

Therefore, if on two different trials a heart rate increase of 10 beats/min is obtained, it does not follow that the two trials are showing an "identical" heart rate response, since the heart rate responses may be reflecting different patterns of neural and/or humoral mediation. Systems theory not only helps us understand this point; it also suggests a way that we can draw differential conclusions regarding underlying mechanisms. The solution is to measure patterns of processes, ideally at different levels, so as to make it possible to test differential interpretations of the data. It should be recalled that a similar point has been made previously with regard to facial EMG (e.g., corrugator activity may be increasing as a function of sadness or concentration; assessing patterns of other muscles allows one to differentiate which state, or combination of states, is being reflected by the observed corrugator activity).

Ax (1953) and Schachter (1957) dealt with this problem at the physiological level by (1) recording multiple channels of information, and (2) scoring each channel in different ways to tap different component processes imbedded in the complex response. For example, from the frontalis muscle region channel, Ax scored the data separately for (1) maximum increase in muscle tension, and (2) number of peaks in muscle tension. Ax found not only that two aspects of "muscle tension" were uncorrelated, but that the maximum muscle tension was significantly higher in anger than in fear, while the number of muscle tensions peaks was significantly higher in fear than in anger.

From the skin conductance channel, Ax scored the data separately for (1) maximum increase skin conductance, and (2) number of rises in skin conductance. Ax found not only that these two aspects of "sweat gland activity" were uncorrelated, but that the maximum increase in skin conductance was significantly higher in fear than in anger, while the number of skin conductance rises was significantly higher in anger than in fear. It seems likely that this pattern of results probably reflects some important set of underlying neuropsychological differences between anger versus fear. However, the physiological interpretation of these patterns remains to be established, and deserves to be pursued in future research.

The important discovery from these early studies was that consistent differences, especially within the cardiovascular system, were found for anger versus fear. Anger was associated with relative increases in peripheral resistance, while fear was associated with relative increases in cardiac output. If any single, easily recordable physiological parameter could be said to tap peripheral resistance, it was diastolic blood pressure. Whereas systolic blood pressure tended to be higher in fear than anger (reflecting increased cardiac output), diastolic blood pressure was significantly higher in anger than fear. In the recent Weerts and Roberts (1976) study, diastolic blood pressure was a major variable distinguishing anger versus fear elicited by imagery.

We (Schwartz *et al.*, 1981) have recently provided an important replication and extension of these earlier findings. Thirty-two college students with a background in acting were instructed on different trials first to imagine, and then to express *nonverbally* while

exercising, one of six different emotional states (happiness, sadness, anger, fear, normal exercise, and relaxation). The exercise task was a modified version of the Harvard step test, which requires subjects to walk up and down a single step.

Systolic and diastolic blood pressure were recorded with an electronic sphygmomanometer, while heart rate was recorded manually by taking the pulse. Two experimenters were used. Both were undergraduate students with no background in physiology, and were naive to the complex hypotheses of the experiment involving patterns of cardiovascular response to the different emotions.

Each trial consisted of two baseline readings taken 1 min apart; one reading taken after the 1-min imagery period (in which subjects *imagined* walking up and down the step, experiencing and expressing the requested emotion); and three readings spaced over approximately 10 min following the 1-min exercise period (in which subjects silently expressed nonverbally the different emotions while they actually walked up and down the step).

The rationale for taking only three relatively simple measures of cardiovascular function was (1) to give the subjects maximum freedom to utilize their bodies both to experience and express the emotions (the prior studies attached many electrodes sensitive to movement artifact that inhibited the subjects' overt behavior in a highly unnatural way; this may in turn have inhibited the magnitude of the cardiovascular patterns evoked in the earlier studies), and (2) to determine whether the findings would be robust enough to be clinically meaningful (and therefore detectable using standard clinical procedures for collecting cardiovascular data).

The rationale for using self-generated imagery followed by exercise was (1) to increase the likelihood that relatively pure emotions would be generated (in the prior studies it is likely that complex blends of anger and fear were evoked, at least in some subjects; also, these studies did not assess the relative emotional purity of their stimulus conditions), and (2) to determine whether allowing subjects to express their emotions overtly would lead to increased physiological patterns that would be clinically meaningful (e.g., it is possible that the style of running in terms of affective expression may have differential consequences for health, with angry running more likely to provoke heart disease and sudden death, and relaxed or happy running more likely to reduce heart disease and sudden death).

In view of the limited reliability of the cardiovascular recording procedures used in this study and the limited number of data points collected, the magnitude and consistency of the results obtained were striking. First, as can be seen in Figure 17-8, different cardiovascular patterns and levels of response were obtained following the imagery period as a function of

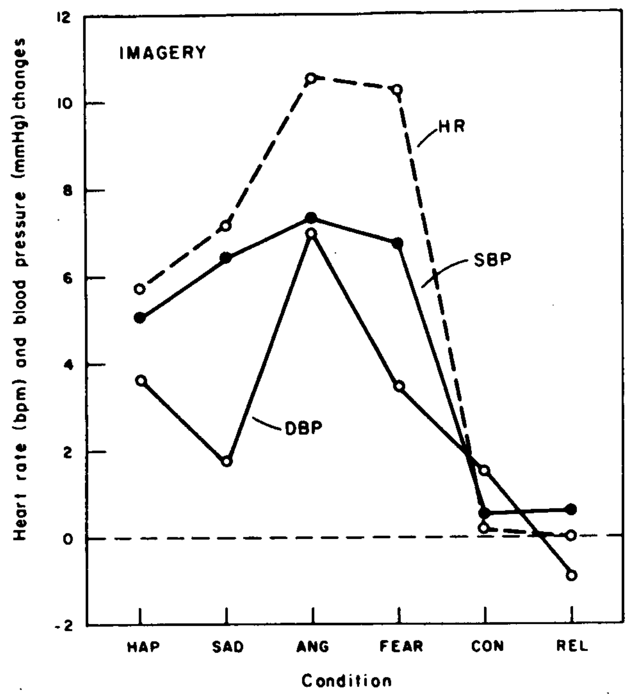


Figure 17-8. Mean changes in heart rate (HR) and in systolic (SB) and diastolic (DBP) blood pressure separately for the happiness (HAP), sadness (SAD), anger (ANG), fear (FEAR), control (CON), and relaxation (REL) conditions following seated imagery.

emotion. The classic finding of diastolic blood pressure being higher in anger than fear was replicated. In addition, both sadness and happiness were differentiated from anger and fear, which in turn were differentiated from control and relaxation.

Following the exercise, large differences in systolic blood pressure and heart rate, but not diastolic blood pressure, were found as a function of the different emotions (see Figure 17-9). Apparently, active exercise produces vasodilation in the muscles and reduces peripheral resistance, which may have overshadowed the relative differences in diastolic pressure between anger and fear. In addition, subjects expressed their anger overtly in this condition. Had subjects been instructed to express anger toward themselves (anger in), perhaps diastolic pressure would have increased after the active exercise. The important point to recognize here is that the cardiovascular patterns in emotion can vary, depending upon the skeletal behavioral state of the individual. Research is now needed that examines the generality of patterns in emotion as a function of different skeletal behavioral states.

Other findings of importance emerged from this study. For example, although systolic blood pressure response immediately following exercise was similar for anger and fear (see Figure 17-9), the *rate of recovery* of systolic blood pressure varied as a function of anger

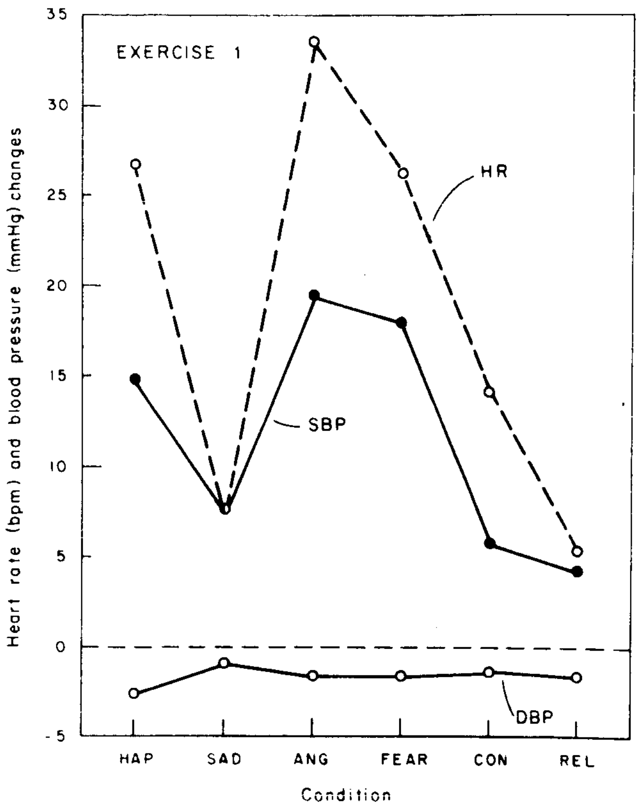


Figure 17-9. Mean changes in heart rate (HR) and in systolic (SBP) and diastolic (DBP) blood pressure separately for the happiness (HAP), sadness (SAD), anger (ANG), fear (FEAR), control (CON), and relaxation (REL) conditions during the first measurement following exercise.

versus fear. Systolic blood pressure was slower to recover following anger than following fear. The hypothesis that different emotions have different "half-lives" is important for basic as well as clinical reasons.

In a recent facial EMG study examining facial muscle patterns in response to elation versus depression self-statements, we (Sirota, Schwartz, & Kristeller, 1983), found that over the course of the experiment, EMG patterns to the elation statements did not grow over time, and EMG levels returned to baselines during rest periods interspersed throughout the experiment. However, EMG patterns to the depression statements grew in intensity over time, and the EMG levels remained high during the rest periods interspersed throughout the experiments. It will be recalled that Izard (1972) proposed that depression was a higher-order emergent emotion reflecting a particular combination of negative emotions, notably sadness and anger. Clearly, future research should record simultaneously patterns of facial muscle and cardiovascular responses as a function of emotion, skeletal behavioral state, and recovery.

Brief mention should be made of the findings obtained when multivariate procedures were applied to the Schwartz *et al.* (1981) data. First, multiple-regression analyses predicting systolic blood pressure from patterns of heart rate and diastolic blood pressure as a function of emotion revealed that the relationship among systolic pressure, diastolic pressure, and heart rate varied as a function of emotion. For example, during imagery, high diastolic pressure is uniquely associated with high systolic pressure during anger, and, in turn, high systolic pressure anger is uniquely associated with *lowered* heart rate. These relations may suggest that during anger the increases in systolic pressure are mediated by increases in peripheral resistance, which in turn may activate inhibition of heart rate through baroreceptor mechanisms. Moreover, discriminant analyses revealed highly significant findings deriving equations that could classify emotional state remarkably correctly as a function of cardiovascular patterning. Although the findings from the pattern classification procedures were rich and informative, space limitations preclude further discussion of these findings here.

As reported elsewhere (Schwartz, 1982), correlations were run among the physiological measures, self-reports of patterns of subjective experience for the imagery and exercise periods, and ratings by the experimenters of patterns of overt emotional expression for the imagery and exercise periods. It turned out that the physiological measures were more strongly and consistently correlated with the observers' judgments than with the subjects' own self-reports! The total set of results did not support the most obvious hypothesis—that the observers may have been using the physiological data unconsciously to make their observational ratings. On the contrary, the findings suggested that the observers were seeing relationships that the subjects themselves did not! For example, observer ratings of fear expression during fear exercise was correlated *negatively* with diastolic blood pressure ($r = .373$, $p < .05$)—a relationship that is highly counterintuitive unless one knows that diastolic pressure typically *decreases* below baseline following isotonic exercise, and that fear should potentiate this effect due to enhanced isotonic exercise. On the other hand, observer ratings of anger expression during an anger exercise were correlated *positively* with diastolic blood pressure ($r = .413$, $p < .05$). Interestingly, self-ratings of fear experience during fear exercise were not correlated with diastolic blood pressure ($r = .01$, n.s.), while self-ratings of anger experience during anger exercise were correlated with diastolic blood pressure ($r = .414$, $p < .05$).

Stimulated by these findings, I (Schwartz, 1982) reviewed the original studies to see whether similar relationships among self-ratings, observer ratings, and physiological measures had been previously examined. Curiously, Schachter (1957) did obtain ob-

server (what he called "expressed") ratings as well as self-reports. Only "mean" (a weighted average of systolic and diastolic) blood pressure correlations were presented in the paper. Schachter found that whereas *self-reports* for both fear and anger were not correlated with mean blood pressure increase, *expressed* behaviors for both fear and anger were significantly correlated with mean blood pressure.

One would hypothesize from a systems perspective that cardiovascular "behavior" and skeletal-motor "behavior" are more intimately connected with each other (both in the periphery and at the level of the brain) than they are connected with the neuropsychological systems involved in monitoring these "behaviors" and making them available to conscious experience. In other words, one's subjective experience includes both the monitoring and interpreting of cardiovascular *and* skeletal-motor processes. It therefore follows that self-report can be more readily dissociated from these two processes than the two processes can be dissociated from themselves. Note that the outside observer "sees" the manifestations of the skeletal-motor "behavior" and then tries to infer from these observations what the person *might* be feeling. In this sense, what the observer does in inferring emotion from overt behavior parallels what a physicist does in inferring the existence of subatomic particles from the "behavior" of bubbles in a cloud chamber: Both are inferences about underlying, organizing processes—an important point, which is returned to at the end of this chapter.

The subject, on the other hand, is not limited in forming and labeling his or her experience solely on the basis of peripheral cues. In fact, people probably vary (among themselves and from situation to situation) with regard to exactly how much they attend to their bodies and how they interpret these cues in forming their experience and self-reports. Because an outside observer is more attentive to overall patterns of overt behavior, an outside observer's ratings are more likely to be consistent with underlying cardiovascular patterns than will the subject's own self-reports.

I return to this issue in the section on personality and the psychophysiology of emotion. The point to emphasize here is the hypothesis that self-report and physiology should be less well connected than physiology is connected with physiology, and that the use of observer ratings can be important in clarifying this issue.

A recent study by Ekman, Levenson, and Friesen (1984) provides additional important support for the hypothesis connecting skeletal-motor behavior—in this case, that of the face—with autonomic patterning in emotion relatively dissociated from subjective experience. Following Schwartz *et al.* (1981), subjects experienced in acting were used (in this case, professional actors, $N = 12$), plus scientists who study the face ($n = 4$). Subjects were instructed to relive six emotions (happiness, sadness, anger, fear, surprise, and disgust), and also to generate overt facial expressions of emotion using instructed movements based on the anatomy of different facial expressions of emotion (Ekman & Friesen, 1978). Heart rate, skin temperature, skin resistance, and forearm flexor muscle tension were recorded. Heart rate was found to differentiate between the positive and negative emotions, while skin temperature further differentiated among the negative emotions. Interestingly, the posed facial muscle movements (which the authors claimed were associated with minimal subjective experience of emotion) led to greater autonomic patterning in emotion than did the relived emotional experiences (which the authors claimed were associated with relatively little facial movement)! The combined findings of Schwartz *et al.* (1981) and Ekman *et al.* (1984) provide important justification of conducting future research that integrates the measurement of skeletal muscle, autonomic indices, and subjective experiences of emotion over time.

CENTRAL NERVOUS SYSTEM PATTERNING AND EMOTION

The degree of subjective, skeletal, and autonomic patterning that is possible depends to a large extent on the degree of patterning of central nervous system processing that is possible. Unfortunately, difficulty in obtaining direct or even indirect electrophysiological measures of localized brain function (through depth electrodes or surface electrodes), coupled with the difficulty in interpreting overt behavior as being an indirect measure of particular neuropsychological processes, has historically led most psychophysiologists interested in the study of emotion to restrict their recording and interpretations to peripheral responses and associated levels of analysis.

However, recent theory and research on hemispheric asymmetry in cognition and emotion have made it possible to raise new questions about cognitive–affective patterning and hemispheric patterning associated with different emotional states. For example, using lateral eye movements as a relative indicator of hemispheric activation, we (Schwartz, Davidson, & Maer, 1975) demonstrated that in right-handed subjects, (1) emotional questions produced relatively more left-eye movements (indicative of right-hemispheric involvement) than nonemotional questions, (2) verbal questions produced relatively more right-eye movements (indicative of left-hemispheric involvement) than spatial questions, and (3) spatial questions produced relatively more stares and blinks than verbal questions. From these three sets of findings, it became possible to uncover discrete patterns of lat-

eral eye movements that could distinguish among all four combinations of cognition and affect: verbal nonemotional, verbal emotional, spatial nonemotional and spatial emotional questions. In other words, not only could affective processes be distinguished from cognitive processes in terms of patterns of eye movement activity, but their interactions could be uncovered as well. It is worth noting that the concept of patterning of cognitive and affective processes at the level of the brain can become a new neuropsychological framework for reinterpreting and extending the original social-psychobiological model proposed by Schachter and Singer (1962).

From a systems perspective, the use of lateral eye movements for the purpose of inferring central nervous system patterning illustrates a changing scientific paradigm regarding the relationship among psychology, physiology, and neurology. As discussed elsewhere (Schwartz, 1978), eye movements can be defined as (1) psychological behavior (if they are simply observed by the naked eye), (2) physiological behavior (if they are recorded on a polygraph), or (3) neurological behavior (if they are interpreted as reflecting underlying neurological processes). The fact that essentially identical findings can be published in different journals reflecting different disciplines is more an indication of the particular conceptual frameworks of the investigators than the actual processes being measured. Interestingly, current research is becoming more "psychoneurophysiological," illustrating the integration and crossing of these three levels.

A major advantage in measuring lateralization of overt behavior and interpreting the findings in neurological terms is that the observations can be made unobtrusively. Thus Sackeim, Gur, and Saucy (1978) have reported that the left side of the face (controlled significantly by the right hemisphere) is more reflective of negative emotions. Their data were based on pictures taken of overt faces and shown to judges who rated left- and right-side composite photographs.

Recently, researchers have attempted to study the emotion–laterality question more closely in terms of fundamental emotions and patterns of self-report. As reviewed in Schwartz, Ahern, and Brown (1979), Tucker (1981), and Davidson (1984), it appears that the hemispheres are differentially lateralized for classes of emotion. A primary hypothesis is that the left hemisphere (in right-handed subjects) is more involved with positive emotions, and the right hemisphere is more involved with negative emotions. For example, we (Schwartz *et al.*, 1979) reported evidence of differential lateralization for positive versus negative emotions in facial EMG recorded from the zygomatic region (which is involved with the smile) while subjects were constructing answers for questions involving positive versus negative emotions: Relatively greater zygomatic facial EMG on the right side of the face for positive emotions was found. A different pattern (relatively greater zygomatic facial EMG on the left side of the face for both positive and negative emotions) was found when subjects were requested to produce voluntarily overt facial expressions of positive and negative emotions.

These findings were replicated and extended (Sirota & Schwartz, 1982) for elation versus depression imagery. The major right versus left zygomatic EMG difference for elation imagery occurred in pure right-handed subjects (right-handed subjects whose parents and siblings were also right-handed). Like the Schwartz *et al.* (1979) results, the Sirota and Schwartz (1982) finding was that voluntarily produced facial expressions did not result in a right-sided increase for positive emotions. As Fridlund and Izard (1983) point out, the interpretation of lateralized facial EMG differences is complex, due to questions of electrode placement, muscle mass, demand characteristics of the situation, and the nature of the emotion task. It is precisely because the facial laterality–emotion hypothesis raises such fundamental methodological and conceptual questions that it is such a fruitful hypothesis for further research.

Lateralized findings for positive versus negative emotions have not been restricted to facial EMG. For example, concerning lateral eye movements, we (Ahern & Schwartz, 1979) reported not only that positive emotions were associated with relatively more right-eye movements than negative emotions, but that these effects were more robust than the previously reported findings for verbal versus spatial processes. We (Ahern & Schwartz, 1979) proposed that the left–right differences in positive versus negative emotions might reflect a more basic difference in approach versus avoidance behavior. We hypothesized that these left–right processes might be mediated subcortically and therefore might be more fundamental than left–right cortical differences in verbal versus spatial processing. A similar conclusion has been proposed by Davidson and colleagues as part of their research program on cerebral laterality and emotion (see Davidson, 1984).

The hypothesis of left–right differences in positive versus negative emotions is most likely oversimplified. Current research is examining patterns of *intra-* as well as *inter*hemispheric processes in different emotions. Davidson, Schwartz, Saron, Bennett, and Goleman (1979) reported findings integrating the two seemingly disparate hypotheses regarding laterality and emotion: (1) that all emotions are lateralized in the right hemisphere, and (2) that emotions are differentially lateralized depending upon their valence (or approach–avoidance tendencies). Using electroencephalogram (EEG) measures recorded from the parietal and frontal regions, they found that parietal EEG showed relatively more activation over the right

hemisphere for *both* positive and negative emotions, whereas frontal EEG showed relatively more activation over the left hemisphere for positive emotions and relatively more activation over the right hemisphere for negative emotions. It is possible that the initial holistic processing of emotional stimuli (a process apparently common to all emotions) may be performed in the right parietal region, whereas the differential interpretation of positive versus negative emotions, and the expression of positive versus negative emotions, are performed by the left and right frontal regions, respectively.

An elegant study by Davidson and Fox (1982) has obtained this overall pattern of EEG findings in two samples of 10-month-old female infants. The infants sat on their mothers' laps while they watched a videotape of an actress generating happy and sad facial expressions. Davidson (1984) has recently proposed that not only are such data consistent with the hypothesis that approach versus avoidance behavior is lateralized in the infant, but moreover that the normal process of development involves the integration of these two different hemispheric styles with the maturation of the corpus callosum.

It follows that differential *intra-* versus *inter*hemispheric patterning should occur for cognitive processes as well as emotional processes. In order to validate the neuropsychological interpretation offered for the earlier lateral eye movement studies (Ahern & Schwartz, 1979; Schwartz *et al.*, 1975) for lateralized patterns of cognitive and affective processes, we (Ahern & Schwartz, 1985) recorded EEG from the frontal and parietal regions while subjects answered affective questions. The EEG was sampled during 4-sec epochs preceding the periods when an eye movement would usually occur (in the study, subjects answered questions with their eyes closed, thus reducing actual eye movements and hence eye movement artifact). Spectral analysis was performed on the data. It was found that for the cognitive dimension, the predicted laterality was found in the posterior (parietal) region (e.g., relatively greater EEG activation for verbal versus spatial questions in the left hemisphere), with little evidence for cognitive lateralization in the anterior (frontal) region. Conversely, for the affective dimension, the predicted laterality was found in the anterior (frontal) region (e.g., relatively greater EEG activation for positive versus negative questions in the left hemisphere), with little evidence for positive-negative lateralization in the anterior (parietal) region. Furthermore, overall relative right posterior (parietal) activation was found for all emotions (replicating Davidson *et al.*, 1979, and Davidson & Fox, 1982).

I return to the question of central nervous system patterning and emotion in the next section on personality and psychophysiological patterning. It seems likely that future research will continue to uncover organized relationships between underlying central nervous system processes and their expression in self-report, physiological activity, and overt behavior. Moreover, more sophisticated psychophysiological techniques such as neuromagnetic measurement promise to open up new vistas for exploring emotion from a biopsychosocial perspective. One challenge will be to keep the technological advances in balance with essential psychosocial advances. Sensitivity to individual differences, instructions, and the social setting will become more and more important as technologically sophisticated research on emotion develops.

INDIVIDUAL DIFFERENCES IN PSYCHOPHYSIOLOGICAL ORGANIZATION AND EMOTION

A major problem and challenge for research on the psychophysiology of emotion involves individual differences in degree of association within psychological and physiological levels and between psychological and physiological levels. The issue of association–dissociation between systems is fundamental to models of disorder within systems (Schwartz,1983), and is often observed in the study of psychopathology. For example, in a clinical setting, Brown and colleagues (Brown, Schwartz, & Sweeney, 1978; Brown, Sweeney, & Schwartz, 1979) have reported that depressed patients and schizophrenic patients differ from each other and from normal controls in how accurately they remember expressing positive affect nonverbally with their faces and bodies. Briefly, compared to observers' ratings of actual overt facial and bodily behavior in a group situation, depressed patients reported experiencing more pleasure than they expressed, whereas schizophrenic patients reported experiencing less pleasure than they expressed.

Dissociations between self-report and behavior, and/or self-report and physiology, are not limited to hospitalized psychiatric patients. Dissociations reliably show up in random samples of relatively healthy college students, and these dissociations have important conceptual, methodogical and clinical implications. A classic example of dissociation between subjective experience and physiological activity and associated overt behavior involves repression. "Repressors" are individuals who have developed the skill of minimizing or avoiding the experience of certain negative emotions. Simply stated, repressors tend to report (and believe) that they are minimally anxious, angry, or depressed, even though their overt behavior and underlying physiology may indicate the opposite.

We (Weinberger, Schwartz, & Davidson, 1979) conducted an experiment to determine whether it was possible to distinguish between people who reported

feeling little anxiety and were accurate (called "true low-anxious"), and people who reported feeling little anxiety but were self-deceptive (called "repressors"). After first splitting subjects on scores on a standard anxiety scale, the low-anxiety-reporting subjects were further split into two subgroups, based on their scores on a second personality scale hypothesized to be sensitive to defensiveness. It turns out that the "social desirability scale" (Crowne & Marlowe, 1964) is not only a measure of social desirability, but also is a reasonably good measure of defensiveness (reviewed in Weinberger *et al.*, 1979). Thus subjects reporting low anxiety were split into a low-defensive/low-anxiety-reporting group (true low-anxiety) and a high-defensive/low-anxiety-reporting group (repressors).

In the experiment, subjects were exposed to a moderately stressful sentence completion task. Subjects were instructed to complete phrases that were neutral, sexual, or aggressive in content. Heart rate, skin resistance, and frontalis EMG from the forehead region were recorded. In addition, subjects' verbal response latencies and measures of verbal disturbance of the subjects' sentence completions were scored. Although there were some interesting patterns observed across measures, the overall findings indicated that repressors generated significantly larger physiological and overt behavioral responses (indicative of negative emotion) than true low-anxiety subjects (even though the repressors actually reported experiencing less anxiety than the true low-anxiety subjects). Furthermore, the magnitude of the repressors' physiological and psychological responses was either equal to, or even greater than, the large-magnitude responses observed in a group of high-anxiety-reporting subjects! These findings have been recently replicated and extended in an important study on "the discrepant repressor" (Asendorph & Scherer, 1985).

The combined findings provide the key for understanding why it has proven so difficult in the past to obtain consistent significant correlations between physiological responses and self-reports across subjects. If a subset of subjects generates erroneous self-report data due to such factors as defensive style (e.g., repression), then not only will correlations across a random sample of subjects below, but the correlations will ultimately be uninterpretable. From a systems point of view, we must distinguish not only among physiological parameters, observer ratings, and self-reports, but we must further distinguish among *different processes that subjects use to label their affective states and the accuracy with which they do so.* Future research on the psychophysiology of emotion must consider individual differences in defensiveness, and must include scales such as the Marlowe–Crowne, if meaningful self-report–physiology relationships are to be uncovered.

I (Schwartz, 1983a) have proposed a general systems theory of disregulation that attempts to explain how systems go out of control. Using the prefix "dis-" across terms, I have proposed that *disattention* (e.g., motivated by a repressive coping style) involves a neuropsychological *disconnection* (to varying degrees), producing a *disregulation* in the system, which is expressed as increased *disorder* in self-regulatory processes (e.g., increased responsivity to stimuli, decreased recovery from stimuli, decreased regularity of rhythms common to homeostatic processes, etc.), which in turn contributes to the development and diagnosis of *disease.* I (Schwartz, 1983a) have reviewed recent data suggesting that individual differences in lateralization to positive versus negative emotions may be related to personality measures of repression and physiological reactivity: Repressors (who may report that things are quite positive, yet express the opposite nonverbally and physiologically) appear to be *relatively functionally disconnected* between the two hemispheres, and thus suffer the consequences of a neuropsychological *disregulation.* Future research on cerebral laterality and emotion should consider the phenomenon of defensiveness, and should include such measures of defensiveness as the Marlowe–Crowne—not only to increase the likelihood of obtaining reliable results, but also the help make more meaningful interpretations of the findings (e.g., the increased laterality in repressors may reflect a conflict between approach and avoidance tendencies, with approach tendencies emphasized by the left hemisphere in right-handed individuals, and avoidance tendencies emphasized by the right hemisphere).

Recent findings (Bowen & Schwartz, in preparation a, in preparation b) provide additional information about disregulation and emotion from a systems point of view. Bowen and I discovered that when subjects are instructed simply to increase their heart rates on some trials and to decrease their heart rates on other trials, subjects vary in the degree to which they respond physiologically in a global undifferentiated (disregulated) fashion versus a more specific, differentiated (self-regulated) fashion. Undifferentiated subjects seem to respond in a rigid manner across situations, suggesting that they emphasize individual stereotypy as described by the Laceys (Lacey & Lacey, 1958). Differentiated subjects seem to respond in a flexible manner across situations, suggesting that they emphasize situational stereotypy as described by the Laceys (Lacey & Lacey, 1958). Using four cardiovascular measures (systolic blood pressure, diastolic blood pressure, heart rate, and pulse transit time), we (Bowen & Schwartz, in preparation a) classified subjects in terms of global cardiovascular arousal (all four measures changed in the same direction in heart rate increase versus decrease trials) versus specific cardiovascular patterning (only one or two measures would change in the same direction in heart rate increase versus decrease trials).

We (Bowen & Schwartz, in preparation a) found

that by splitting subjects into undifferentiated and differentiated groups, it was possible to predict which subjects would show cardiovascular differentiation to different emotions. The undifferentiated subjects responded with overall cardiovascular arousal to imagined positive and negative emotions, whereas the differentiated subjects responded to the different imagined positive and negative emotions with differentiated patterns of cardiovascular activity. The undifferentiated subjects generated self-reports, particularly of happiness and anger, indicating a very simple, stereotypic emotional experience, whereas the differentiated subjects gave self-reports indicating a more complex, rich, blended set of emotional experiences.

In a replication and extension of these findings, we (Bowen & Schwartz, in preparation b) repeated the study using an independent sample of subjects, this time including measures of defensiveness (the Marlowe–Crowne), laterality (facial EMG), and health (self-reports of illness). Not only were the original findings replicated, but in addition the undifferentiated subjects were found to (1) score significantly higher on the Marlowe–Crowne, (2) to show evidence of increased laterality in facial EMG, and (3) to report increased illnesses.

This pattern of findings is not only consistent with the repression–cerebral disconnection–disease hypothesis (Schwartz, 1983a), but may also be related to other disattention syndromes, such as Type A behavior (discussed in Schwartz, 1983b). The important point to recognize here is that discrepancies between self-reports and physiological responses, when interpreted through the perspective of systems theory, become particularly rich and important sources of data in their own right. Patterns of discrepancies can have important implications for theory, for research, and possibly for clinical practice. Future research exploring physiological–subjective relationships in emotion will probably profit from looking closely at individual differences within and across levels of patterns of processes.

EMOTION AS BIOPSYCHOSOCIAL ORGANIZATION: METHODOLOGICAL IMPLICATIONS

The hypothesis that the concept of emotion implies not only a set of feeling states, physiological reactions, motivational expressions, and behaviors, but an *organization* of these processes to meet specific biopsychosocial goals, is a fundamental application of systems theory. A focus on organization leads us to focus our attention on the search for replicable patterns of processes within and across levels. These patterns, as indicated in the preceding section, can vary in their complexity and stability as a function of individual differences. Viewing the individual difference variation from the perspective of levels and complexity of organization has the potential to integrate research on the psychophysiology of emotion with research on personality and psychopathology.

A focus on patterns of processes implies more than just systematically assessing patterns of subjective experience (e.g., using instruments such as the DES developed by Izard, 1972), patterns of physiological responses, or patterns of overt behavioral expression in a social context. It implies that we develop more sophisticated and meaningful ways for statistically revealing the underlying organization that is present. Multivariate statistics have been usefully applied to cardiovascular (Schwartz *et al.*, 1981) and facial EMG (Fridlund *et al.*, 1984) data, and it seems likely that future advances in mathematics and statistics, particularly as developed in cognitive science, artificial intelligence, and robotics research will find significant applications to future research on the psychophysiology of emotion.

Generally speaking, as noted by various authors (e.g., Fridlund & Izard, 1983; Schwartz, 1982), the concept of organization encourages one to look more precisely at psychophysiological patterns in specific stimulus–response configurations. For example, if subjects are watching a film, and are generating different facial expressions as the film unfolds, it seems prudent to examine psychophysiological patterns *as they are organized at the precise moments when particular facial expressions of emotion occur* (e.g., within a few seconds, as opposed to simply averaging all this information over minutes, or sampling responses in fixed time without regard to the flow of behavior over time). Emphasis on organization in systems terms encourages us to look for organization in meaningful biopsychosocial contexts and durations. This is clearly a challenge for future research.

Another methodological consideration clarified by a systems approach to emotion concerns the role that social variables play in the psychophysiological patterns observed in a given situation. Not only do electrodes constrain movement (and therefore emotional expression), but implicit if not explicit instructions to refrain from moving may alter the meaningfulness of the data obtained (recall that in Schwartz *et al.*, 1981, even though patterns of cardiovascular response were found to vary in emotion following both imagery and overt exercise conditions, the *organization* of the patterns was different in the imagery and exercise conditions). The use of video cameras, the nature of the instructions used for obtaining self-reports of emotion, and the amount of self-report information sampled all have the potential to alter the psychophysiological organization obtained. This realization clearly makes research on emotion more complicated and more challenging, but the challenge can be met successfully if a biopsychosocial view of emotion is

kept clearly in mind, and care is taken to design research from the perspective of biopsychosocial measurement.

The general pattern classification approach of Fridlund *et al.* (1984) is instructive, in that it emphasizes feature extraction as an important component of pattern analysis. From a systems perspective, "single" measures such as "corrugator EMG" are really "multiple" measures in the sense that biological signals carry complex patterns of information that can be extracted. The classic study by Ax (1953), as discussed previously, illustrates this general principle by demonstrating that it is possible to differentiate fear from anger *within a "single" response* (e.g., comparing maximum increase in skin conductance with number of rises in skin conductance). According to systems theory, it is possible to have patterning *within* individual physiological measures, since all "wholes" represent organized patterns of "parts." The recent work by Cacioppo and colleagues applying this kind of methodology to facial EMG in cognition and emotion is an important advance in this regard (Cacioppo & Petty, 1983).

EMOTION AS BIOPSYCHOSOCIAL ORGANIZATION: CONCEPTUAL IMPLICATIONS

The hypothesis that emotion reflects biopsychosocial organization has important conceptual implications that can be operationalized and put to empirical tests. One example involves the hypothesis that emotion is revealed as an emergent property of multiple interacting systems within and across levels. According to systems theory, the subjective experience of emotion should be more stable and more complete as more physiological elements are activated and organized in meaningful patterns. An approach to studying this question is to use biofeedback as a methodology for producing different combinations and patterns of physiological responses, and for examining the subjective changes that covary with the physiological patterns. For example, when subjects are taught to increase both their heart rate and frontalis muscle tension simultaneously, they report experiencing more anxiety than if they increase heart rate alone or frontalis muscle tension alone (see Schwartz, 1977).

An excellent chapter by Leventhal (1980) proposes some new aspects of emotion that are consistent with the emergent principle. Leventhal proposes that although patterns of bodily feedback contribute to the emergent experience of emotion, the emotional experience may be disrupted (if not destroyed) if subjects are instructed to attend voluntarily to *specific bodily parts*. This disruptive effect is predicted from systems theory. Attending to a subset of parts removes information from certain processes and alters others. Therefore, focusing one's attention can attenuate, if not eliminate, certain emergent properties that depend upon the interaction of the multiple components for their existence. Focused attention acting as a "disemergent" may be a mechanism used by repressors to dampen their emotional experiences. An analogy would be how the perception of a forest can be disrupted or destroyed if one attends specifically to the trees.

Another implication of a systems approach to emotion involves levels of organization and the development of organization as applied to individual differences in emotional experience and expression. Our recent research (Bowen & Schwartz, in preparation a, and in preparation b) distinguishing between undifferentiated (rigid) and differentiated (flexible) cardiovascular responders suggests that subjects do vary in their ability to generate differentiated physiological and psychological patterns to different emotions. Lane and I (Lane & Schwartz, in preparation) have proposed that stages of cognitive development and stages of emotional development generally unfold in parallel and are organized by the frontal and prefrontal cortex. We have hypothesized that with increased emotional development, there is increased capacity for differentiation in biological, psychological, and social levels of functioning. The conflict in the psychophysiology literature between "arousal" versus "pattern" theorists may be caused in part by differences in subject populations sampled, who may have varied in their levels of cognitive and emotion development, and therefore in their physiological and social development as well. A challenge for future research is to view emotion from a developmental perspective (Davidson & Fox, 1984) and then to capture individual differences in the capacity to differentiate (and integrate) higher levels of physiological and psychological organization—from undifferentiated globality (e.g., being "upset") to differentiated specificity (e.g., being "disappointed"). We (Lane & Schwartz, in preparation) have proposed that a systems approach to organization and complexity may improve our capacity to understand the relationship among the psychophysiology of emotion, psychopathology, and physical disease.

This chapter has not specifically dealt with the relationship between emotion and cognition. However, since a systems approach to emotion has implications for the emotion–cognition relationship, a brief comment about this fundamental question is worth making here. As mentioned in the "Introduction and Overview," the concept of emotion is ultimately an inferred concept, not unlike inferred concepts from modern physics. There is a curious and, I believe, an important parallel between the challenges facing modern physics and those facing modern psychophysiology. In modern physics, scientists observe the behavior of meters or graphs generated by electronic

machines, or the path of bubbles in cloud chambers, and attempt to infer underlying particles or forces to explain the organization or pattern of behavior observed. In psychophysiology, scientists observe the behavior of meters or graphs generated by electronic machines, or the path of behavior on a video screen, and attempt to infer underlying "particles" (thoughts) or "forces" (emotions) to explain the organization or pattern of behavior observed. Psychophysiology has one advantage over modern physics, in that its subjects can attempt to describe what they are thinking and feeling verbally (thereby providing another set of observations). But ultimately, the notion of inference becomes essential for understanding the way the science is practiced and evolves.

I believe that some of the conceptual difficulties facing modern physics have parallels in modern psychophysiology, and that we should reflect upon some of the potential implications of these parallels, since the parallels may themselves reflect systems metaprinciples (see Schwartz, 1984). For example, most behavioral and biomedical researchers view cognition and emotion as two separate, yet interacting processes, and researchers have attempted to classify some parts of the brain as more "emotional" (e.g., limbic structures) and other parts as more "cognitive" (e.g., cortical structures). However, as Zajonc (1980) has recently pointed out, all cognitive processes have affective components, and all emotional processes have cognitive components. An alternative view, one suggested by modern physics, is the hypothesis that cognition and emotion reflect two different qualities or aspects of a whole that has yet to be labeled. For example, it is now well established that light has both wave-like and particle-like properties. If an experiment is set up to measure wave properties, light will appear to function as a wave, whereas if an experiment is set up to measure particle properties, light will appear to function as a particle. However, it is difficult to conceptualize light as being both a wave and particle, shifting its relative emphasis back and forth. The concept of a "wavicle" is sometimes used to express the idea that light is not a wave versus a particle, but instead is a whole that includes wave-like and particle-like properties.

It is possible that "emotion may be to cognition as waves are to particles." In other words, it is possible that emotion and cognition are two qualities of a whole, which for lack of a better term might be called "cogmotion" (Schwartz, 1984). All levels of functioning in the nervous system from the brain stem through the prefrontal cortex, may have degrees of functioning that reflect both cognitive and affective qualities of functioning. Experiments that focus more on the cognitive versus affective qualities of functioning may emphasize and reveal the cognitive versus affective qualities of human functioning, altering the system in ways predicted by Heisenberg's uncertainty principle (which says, in essence, that if you attempt to measure one thing, not only does this influence the thing you are measuring, but it makes it difficult to assess other things because of built-in uncertainty).

Since systems theory encourages us to think about parallels between and among all levels and disciplines (including the subatomic), the reader should consider the parallel proposed here between wave-particle theory and emotion-cognition theory as a general analogy whose purpose is to stimulate new ways of thinking about the relationship between emotion and cognition, and therefore about new ways of designing experiments and interpreting data. If "cogmotion" is like "light" in the sense that the terms "cognition" and "emotion" may refer to two different qualities of the organized functioning of organisms, the relationship between thoughts and feelings may be more intimate than heretofore conceived. Psychophysiology may have the potential to uncover the implicit organization inherent in cognitive–affective integration, and thereby to provide a new window for connecting theory and research on cognition with theory and research on emotion.

ACKNOWLEDGMENTS

Preparation of this chapter was supported by a grant from the National Science Foundation. The chapter reflects an extension of Schwartz (1982).

REFERENCES

Ahern, G. L. & Schwartz, G. E. Differential lateralization for positive versus negative emotion. *Neuropsychologia*, 1979, 17, 693–697.

Ahern, G. L. & Schwartz, G. E. *Differential lateralization for positive versus negative emotion in the human brain: EEG spectral analysis.* Manuscript in preparation.*

Asendorph, J. B., & Scherer, K. R. The discrepant repressor: differentiation between low anxiety, high anxiety and repression of anxiety by autonomic-facial-verbal patterns of behavior. *Journal of Personality and Social Psychology*, 1985, 45, 1334–1346.

Ax, A. F. The physiological differentiation between fear and anger in humans. *Psychosomatic Medicine*, 1953, 15, 433–442.

Basmajian, V. J. *Muscles alive: Their functions revealed by electromyography* (4th ed.). Baltimore: Williams & Wilkins, 1978.

Bowen, W., & Schwartz, G. E. Self regulated rigid & flexible cardiovascular responding: Relationship to emotion. Manuscript in preparation. (a)

Bowen, W., & Schwartz, G. E. Self regulated rigid & flexible cardiovascular responding: relationship to emotion, repression and health. Manuscript in preparation. (b)

Brown, S. L., Schwartz, G. E., & Sweeney, D. R. Dissociation of self-reported and observed pleasure in depression. *Psychosomatic Medicine*, 1978, 40, 536–548.

Brown, S. L., Sweeney, D. R., & Schwartz, G. E. Differences between self-reported and observed pleasure in depression and schizophrenia. *Journal of Nervous & Mental Disease*, 1979, 167, 410–415.

Cacioppo, J. T., & Petty, R. E. (Eds.). *Social psychophysiology*. New York: Guilford Press, 1983.

Cacioppo, J. T., & Petty, R. E. Electromyograms as measures of

* Neuropsychologia, 1985, vol. 23 pp. 745-755.

extent and affectivity of information processing. *American Psychologist*, 1981, 36, 441–456.
Crowne, D. P., & Marlowe, D. *The approval motive: Studies in evaluative dependence.* New York: Wiley, 1964.
Darwin, C. *The expression of the emotions in man and animals.* Chicago: University of Chicago Press, 1965. (Originally published 1872.)
Davidson, R. J., Schwartz, G. E., Saron, C., Bennett, J., & Goleman, D. J. Frontal versus parietal EEG asymmetry during positive and negative affect. *Psychophysiology*, 1979, 16, 202–203.
Davidson, R. J. Affective, cognition and hemispheric specialization. In C. E. Izard, J. Kagan, & R. Zajonc (Eds.), *Emotion, cognition and behavior.* New York & London: Cambridge University Press, 1984.
Davidson, R. J., & Fox, N. A. Asymmetrical brain activity discriminates between positive versus negative affective stimuli in 10-month-old human infants. *Science*, 1982, 218, 1235–1236.
Davidson, R. J., & Fox, N. A. *The psychobiology of affective development.* Hillsdale, NJ: Erlbaum, 1984.
deRosnay, J. *The macroscope.* New York: Harper & Row, 1979.
Ekman, P., Friesen, W. V., & Ellsworth, P. C. *Emotion in the human face.* Elmsford, NY: Pergamon Press, 1972.
Ekman, P., & Friesen, W. V. *Unmasking the face.* Englewood Cliffs, NJ: Prentice Hall, 1975.*
Ekman, P., & Friesen, W. V. *Facial action coding system* (FACS). Palo Alto: Consulting Psychologists Press, 1978.
Ekman, P., Levenson, R., & Friesen, W. V. Autonomic nervous system activity distinguishes among emotions. *Science*, 1984, 221, 1208–1210.
Engel, G. L. The need for a new medical model: a challenge for biomedicine. *Science*, 1977, 196, 129–136.
Fridlund, A. J., & Izard, C. E. Electromyographic studies of facial expressions of emotions. In J. T. Cacioppo & R. E. Petty (Eds.) *Social psychophysiology.* New York: Guilford Press, 1983.
Fridlund, A. J., Schwartz, G. E., & Fowler, S. C. Pattern recognition of self-reported states from multiple site facial EMG activity during affective imagery. *Psychophysiology*, 1984, 21, 622–637.
Greenfield, N. S., & Sternbach, R. A. (Eds.). *Handbook of psychophysiology.* New York: Holt, 1972.
Izard, C. E. *The face of emotion.* New York: Appleton-Century-Crofts, 1971.
Izard, C. E. *Patterns of emotions.* New York: Academic Press, 1972.
Izard, C. E. *Human emotions.* New York: Plenum Press, 1977.
James, W. What is an emotion. *Mind*, 1884, 9, 188–204.
Lacey, J. I., & Lacey, B. C. Verification and extension of the principle of autonomic response stereotyping. *American Journal of Psychology*, 1958, 71, 50–73.
Lane, R. D., & Schwartz, G. E. *A systems theory of emotional development: Integration of the biological, psychological and social perspective.* Manuscript in preparation.+
Lang, P. J. A bio-informational theory of emotional imagery. *Psychophysiology*, 1979, 16, 495–512.
Lang, P. J., Miller, G. A., & Levin, D. N. Anxiety and fear: Central processing and peripheral physiology. In R. J. Davidson, G. E. Schwartz, and D. Shapiro (Eds.), *Consciousness and Self-Regulation: Advances in Research, Volume 3* (pp. 123–151). New York: Plenum, 1983.
Lang, P. J., Rice, D. G., & Sternbach, R. A. Psychophysiology of emotion. In N. S. Greenfield & R. A. Sternbach (Eds.), *Handbook of psychophysiology.* New York: Holt, 1972.
Leigh, H., Reiser, M. F. *The patient: Biological, psychological and social dimensions of medical practice.* New York: Plenum Press, 1980.
Leventhal, H. Toward a comprehensive theory of emotion. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* New York: Academic Press, 1980.
McGuigan, F. J. Imagery and thinking: Covert functioning of the motor system. In G. E. Schwartz & D. Shapiro (Eds) *Consciousness and self-regulation: Advances in research and theory* (Vol. 2). New York: Plenum Press, 1978.
Miller, J. G. *Living systems.* New York: McGraw-Hill, 1978.
Phillips, D. C. *Holistic thought in social science.* Stanford University Press, 1976.
Plutchik, R. *Emotion: A psychoevolutionary synthesis.* New York: Harper & Row, 1980.
Polonsky, W. H., & Schwartz, G. E. *Facial electromyography and the self-deceptive coping style. Individual differences in the hemispheric lateralization of affect.* Manuscript submitted for publication, 1984.
Sackeim, H. A., Gur, R. C., & Saucy, M. C. Emotions are expressed more intensely on the left side of the face. *Science*, 1978, 202, 434–435.
Schachter, J. Pain, fear and anger in hypertensive and normotensives. *Psychosomatic Medicine*, 1957, 19, 17–29.
Schachter, J., & Singer, J. E. Cognitive, social and physiological determinants of emotional state. *Psychological Review*, 1962, 69, 379–399.
Schwartz, G. E. Biofeedback, self-regulation and the patterning of physiological processes. *American Scientist*, 1975, 63,314–324.*
Schwartz, G. E. Psychosomatic disorders and biofeedback: A psychobiological model of disregulation. In J. D. Maser & M. E. P. Seligman (Eds.), *Psychopathology: Theory and practice.* Baltimore: Williams & Wilkins, 1977.
Schwartz, G. E. Psychobiological foundations of psychotherapy and behavior change. In S. L. Garfield & A. E. Bergin (Eds.) *Handbook of psychotherapy and behavior change* (2nd ed.). New York: Wiley, 1978.
Schwartz, G. E. The brain as a health care system. In G. Stone, N. Adler, & F. Cohen (Eds.), *Health psychology.* San Francisco: Jossey-Bass, 1979.*
Schwartz, G. E. Behavioral medicine and systems theory: A new synthesis. *National Forum*, 1980, 4, 25–30.
Schwartz, G. E. A systems analysis of psychobiology and behavior therapy: Implications for behavioral medicine. In H. Leigh (Ed.), *Psychotherapy and Psychosomatics*, 36, Special issue on behavioral medicine, 1981, 159–184.
Schwartz, G. E. Cardiovascular psychophysiology: A systems perspective. In J. T. Cacioppo & R. E. Petty (Eds.), *Focus on cardiovascular psychophysiaology.* New York: Guilford Press, 1982.
Schwartz, G. E. Disregulation theory and disease: Applications to the repression/cerebral disconnection/cardiovascular disorder hypothesis. (Special edition). *International Review of Applied Psychology*, 1983, 32, 95–118. (a)
Schwartz, G. E. Social psychophysiology and behavioral medicine: A systems perspective. In J. T. Cacioppo & R. E. Petty (Eds.), *Social psychophysiology.* New York: Guilford Press, 1983. (b)
Schwartz, G. E. Psychobiology of health: A new synthesis. In B. L. Hammonds & C. J. Scheirer (Eds.), *Psychology and health, The Master Lecture Series* (Vol. 3). Washington: American Psychological Association, 1984.
Schwartz, G. E., Ahern, G. L., & Brown, S. L. Lateralized facial muscle response to positive versus negative emotional stimuli. *Psychophysiology*, 1979, 16, 561–571.
Schwartz, G. E., & Beatty, J. (Eds.), *Biofeedback: Theory and research.* New York: Academic Press, 1977.
Schwartz, G. E., Brown, S. L., & Ahern, G. L. Facial muscle patterning and subjective experience during affective imagery: Sex differences. *Psychophysiology*, 1980, 17, 75–82.
Schwartz, G. E., Davidson, R. J., & Maer, F. Right hemisphere lateralization for emotion in the human brain: Interaction with cognition. *Science*, 1975, 190, 286–288.
Schwartz, G. E., Fair, P. L., Salt, P., Mandel, M. R., & Klerman, G. L. Facial muscle patterning to affective imagery in depressed and nondepressed subjects. *Science*, 1976, 192, 489–491. (a)
Schwartz, G. E., Fair, P. L., Salt, P., Mandel, M. R., & Klerman, G. L. Facial expression and imagery in depression: An electromyographic study. *Psychosomatic Medicine*, 1976, 38, 337–347. (b)
Schwartz, G. E., & Weinberger, D. A. Patterns of emotional re-

* Not referenced in this chapter.
+ Levels of emotional awareness: A cognitive-developmental theory and its application to psychopathology. Amer. J. of Psychiatry, 1987, vol. 144(2), pp. 133-143.

sponses to affective situations: Relations among happiness, sadness, anger, fear, depression and anxiety. *Motivation and Emotion*, 1980, 4, 175-191.

Schwartz, G. E., Weinberger, D. A., & Singer, J. A. Cardiovascular differentiation of happiness, sadness, anger and fear following imagery and exercise. *Psychosomatic Medicine*, 1981, 43, 343-364.

Sirota, A. D., & Schwartz, G. E. Facial muscle patterning and lateralization during elation and depression imagery. *Journal of Abnormal Psychology*, 1982, 91, 25-34.

Sirota, A. D., Schwartz, G. E., & Kristeller, J. Facial muscle activity during induced mood states: differential growth and carry over of elated versus depressed patterns. *Psychophysiology*, 1983, 20, 471.

Sternberg, R. Intelligence, information processing and analogical reasoning: The componential analysis of human abilities. Hillsdale, NJ: Erlbaum, 1977.

Tompkins, S. S. Affect as amplification: Some modifications in theory. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research and experience*. New York: Academic Press, 1980.

Tucker, D. M. Lateral brain function, emotion and conceptualization. *Psychological Bulletin*, 1981, 84, 19-46.

von Bertalanffy, L. *General systems theory*. New York: Braziller, 1968.

Weerts, T. C., & Roberts, R. The physiological effects of imagining anger provoking and fear provoking scenes. *Psychophysiology*, 1976, 13, 174.

Weinberger, D. A., Schwartz, G. E., & Davidson, R. J. Low anxious, high anxious and repressive coping styles: Psychometric patterns and behavioral and physiological responses to stress. *Journal of Abnormal Psychology*, 1979, 88, 369-380.

Zajonc, R. B. Feeling and thinking: Preferences need no interferences. *American Psychologist*, 1980, 35, 151-175.

5,-53 N88-23371 140789 18P.

AN OVERVIEW OF CURRENT APPROACHES AND FUTURE CHALLENGES IN PHYSIOLOGICAL MONITORING

Richard L. Horst
ARD Corporation
Columbia, MD

ABSTRACT

Sufficient evidence exists from laboratory studies to suggest that physiological measures can be useful as an adjunct to behavioral and subjective measures of human performance and capabilities. Thus it is reasonable to address the conceptual and engineering challenges that arise in applying this technology in operational settings. The present paper will attempt to identify such application-oriented issues and to provide an overview of the state-of-this-art. Issues to be reviewed will include the advantages and disadvantages of constructs such as mental states, the need for physiological measures of performance, areas of application for physiological measures in operational settings, which measures appear to be most useful, problem areas that arise in the use of these measures in operational settings, and directions for future development.

INTRODUCTION

Prospects for the routine use of physiological monitoring in operational settings are becoming more favorable. This situation is due in part to advances in recording technology, in part to research results that suggest the usefulness of physiological data, and in part to an increasingly critical perceived need for information about the status of the human operator in complex man-machine systems.

One can sometimes gain an impression of the state of one's art by the criticism it receives during informal exchanges. Not many years ago, those of us involved in psychophysiological research, and in particular scalp-recorded brain-wave measurement, were frequently asked to endure two comments:

> "Surface recordings provide only a gross indication of brain function. It's like putting an electrode on the outside of a computer and trying to infer the processes going on inside."

and

> "How can you interpret these field potential phenomena without understanding the underlying mechanisms, if not the underlying physiology?"

Perhaps it is the company one keeps, but lately other comments have been heard more frequently:

> "You can't have electrode wires dangling from a pilot in the cockpit."

> "Operators will never accept having their physiology monitored. It takes too long to hook them up. It's too messy. Besides, pilots will be afraid that you'll turn up some arrhythmia that could ground them."

"What do you do with all the electrical artifacts that are likely to show up in operational settings? In the laboratory you can reject contaminated data and keep collecting until you get enough clean data. In the field you will not have that luxury."

"There is no one-to-one relationship between (fill in your favorite physiological sign) and performance. You would have to know a lot about overt behavior in order to interpret concurrently recorded physiological measures. And if you have the behavioral measures, why do you need the physiological?"

Thus, the issues of concern seem to be changing, from questioning the basic value of the measures to questioning how one implements them in applied settings. There is no question that much basic research and theorizing remain to be done in this field. We don't yet have a good understanding of the functional significance of many psychophysiological phenomena. But, as funding permits, progress is being made and physiological measures are proving to be valuable adjuncts to behavioral and subjective measures in the assessment of human performance (see Ref. 1 for a recent broad survey of this field). For this purpose, derived measures of physiological signals can be useful as dependent measures, regardless of how poorly we understand the underlying physiology. A thorough understanding of source generator loci and cellular mechanisms would, no doubt, enhance the interpretive power of these measures; but as long as they vary systematically with experimental manipulations, these indices can be used, as are behavioral and subjective measures, in the monitoring, prediction, and diagnosis of performance.

Corresponding to this shift in the concerns of critics, one notices an attitudinal change among practitioners. For years, basic researchers took a rather cavalier approach -- that their role was to demonstrate the value of psychophysiological measures of performance and to uncover the relationships between these measures and conceptual information-processing constructs. Problems related to the transition of this technology to applied task environments and the implementation of these measures in the field could be left to "the engineers." Now, one finds considerable interest, among both researchers and funding agencies (one can speculate about the causal relationships here), in beginning to address these deferred "engineering" problems. Impetus has been provided by advances in a number of enabling technologies — micro-electronics, signal processing, wireless communications, display technology, and artificial intelligence (AI). Consequently, laboratory work is being conducted with an eye towards task scenarios and measurement protocols that could, with modification, be used in the field. More research, both basic and applied, is being conducted in simulators.

All of this represents progress and suggests the need to look closely at the realistic prospects for applying physiological measures in operational settings. The remainder of this paper will provide a necessarily brief overview of some of these prospects, the approaches that are currently being pursued, the state-of-the-art, and recommendations for future directions in research and development. One theme, which corresponds to the topic of this workshop, will be the prospects for quantifying operator mental states.

MENTAL STATE ESTIMATION

It is interesting that in the conceptual plans for such next-generation

systems as those involving Super-Cockpits, one sees a recognition of the fact that operator mental status is something the system should measure and to which it must adapt. No doubt, this design goal follows from the recognition that, under some operational scenarios, the human operator could be the limiting factor for successful mission completion. These systems will be capable of presenting more information than even a fully functional human can process, and some of the threats faced in the operational environment, e.g., high G load or chemical/biological/radialogical (CBR) agents, could disable the operator without fatally impairing system hardware and software. Moreover, these systems are expected to have sufficient automated subsystems and artificial intelligence that the system could aid an overburdened operator or, to some extent, take over for an impaired operator.

Certainly, therefore, the ability to assess the functional mental status of the human operator is of critical importance in these systems, and would be useful to the designers of many less exotic systems. But how far can we take this concept? Can one conceptualize functional mental status in terms of a finite number of discrete mental states? Is there some value to being able to classify the human operator from moment to moment as being in a state of high or low workload, fatigue, boredom, confusion, stress, or any of the numerous other explanatory constructs that we invoke, even informally, in interpreting our data or in designing our man-machine interfaces?

Typically, these constructs are operationally defined in terms of experimental variables. Beyond that, it is not yet clear whether such discrete states exist, or with what taxonomy they should be classified. Operator effectiveness is ultimately defined in terms of behavioral output. However, there seems to be both diagnostic and prescriptive value in attempting to develop such a taxonomy of mental constructs, rather than focusing just on observable task performance. For example, task performance may deteriorate for a wide variety of reasons. An operator may miss an alarm signal either because he was cognitively overloaded or because he was bored and not sufficiently vigilant. A system designer, or co-pilot, would take different remedial actions, depending on which of these "states" led to the degradation in performance. Furthermore, many task environments allow the human operator to function with some spare capacity such that, to some extent, increased task demands can be met with increased effort in order to maintain behavioral output at a relatively constant level. In such situations, mental state indices may predict susceptibility to an impending deterioration in performance, should task demands increase still further. Finally, when task demands are low, there may be little behavioral output from which one can gauge the status of the operator. A sense of the operator's mental state in such situations could be used to infer whether or not such lack of responding was appropriate and the extent to which the operator is prepared to respond appropriately should conditions change. Therefore, the diagnostic and, hopefully, prescriptive value of mental state constructs are somewhat akin to that of clinical syndromes. Analogous to the different treatments which may be prescribed depending on a clinical diagnosis, inferences about the mental states which underlie an observed performance deficit may suggest alternative design or operational "treatments."

The danger in using mental state conceptualizations to explain data, of course, lies in our tendency to think that if we can label something, we have understood it. Terms like "boredom" may not imply the same "syndrome" to everyone. Therefore, until we have sufficient data to define what are the

distinguishing features and performance-related consequences of "boredom," it is imperative that we continue to operationally define our use of such terms.

THE VALUE OF PHYSIOLOGICAL MEASURES

Regardless of the stock one puts in the explanatory power of mental states, it follows from the above discussion that it would be unwise to evaluate and predict an operator's ability to perform solely from observing behavior on a primary task. Performance on secondary tasks can be instructive for measuring the processing capacity entailed by a primary task. However, with this approach it is difficult to ensure that the operator always gives mental priority to the primary task, the results may be of questionable validity if used to generalize to situations in which the primary task is performed alone, and incompatibilities between the behavioral responses required by the two tasks may make it difficult to draw inferences about the demands placed on perceptual or decision-making processes. Moreover, the sort of contrived secondary tasks that have often been used in laboratory studies are clearly not acceptable in operational settings, so secondary task measures must be found among the activities that the operator is doing in the course of normal operations.

Simply asking the operator for subjective ratings of his perceived state is often useful, but is also fraught with difficulties. The operator may not realize that his environmentally-defined workload is high when, in fact, it is. Furthermore, such subjective ratings tend to be unreliable when administered in operational settings while the operator is simultaneously trying to maintain task performance, and the mere act of completing the rating itself, of course, constitutes an additional task burden on the operator.

For these reasons, there is considerable appeal to the prospects of gaining additional information about the functional status of operators from their physiological signs. As discussed later in this paper, much evidence now suggests that, if interpreted in conjunction with behavioral and subjective measures, physiological indices offer the possibility of objectively inferring, not only the general physical fitness to perform, but also the cognitive status of an operator. Physiological measures can often be used to confirm the conclusions derived from behavioral or subjective measures. There are also instances in the literature of physiological measures providing complementary information regarding cognitive activity to that which is available from behavioral measures.

While there is a certain intuitive appeal to the objectivity and non-intrusiveness afforded by physiological measures of mental processes, the possible limitations of this technology have been pointed out by a number of critics. Johnson (Ref. 2) has listed several typical concerns:

- Most research studies have used performance changes to interpret physiological changes; it is the inverse problem, using physiological indices to predict performance, that is of interest in operational settings, and most attempts to take this approach have been disappointing.

- There are not specific physiological response patterns associated with specific behaviors or specific states; task difficulty plays an important mediating role.

- There are large individual differences in physiological responses; response differences due to individual response stereotypy tend to be larger than differences due to situational response stereotypy.

Zacharias (Ref. 3) has likewise faulted most physiological work for failing to take account of the effects of task difficulty on the measures of interest. He also points out that while attempts to more fully characterize physiological status by creating a vector of physiological indices may provide increased correlations with, for example, measures of workload, there can actually be a reduction in the statistical significance of such correlations, as "an increasing number of noisy physiologic indicators are included in the actuation vector."

While these criticisms are well-taken and must be addressed by those wishing to use physiological measures of performance, they pose no insurmountable problems for the knowledgeable application of physiological monitoring technology. It is possible to deal with, and in fact take advantage of, the manner in which physiological indices reflect task difficulty (see, for example, Samaras'[1] paper in the present Proceedings). The irrefutable fact that individual differences exist, may likewise be turned to our advantage. In most operational settings we are dealing with highly trained operators, and it is technologically possible to customize the parameters of a monitoring system for the individual operator. Finally, the question of whether or not unique configurations of physiological patterns can be associated with particular mental states may be moot, if one assumes that interpretations can be based on <u>changes</u> in physiological indices viewed in conjunction with <u>changes</u> in operator behavior or system performance. In other words, one rarely would be faced with the need to classify operator state in an absolute sense. The more frequent, and more manageable, challenge would be to classify <u>changes</u> in state or functional status, in relative terms, with reference to task performance and other behavioral data.

AREAS OF APPLICATION

Physiological measures can be useful in operational settings for a variety of purposes. Other papers in this session have presented some specific operational settings of interest. Most uses can be seen to fall into one of the following categories:

<u>System Design</u>. Reducing operator workload and drawing an operator's attention to certain task-related stimuli are often design goals. To the extent that physiological measures are reliable indices of these mental constructs, they can be used to make design decisions. For this group of applications, recording in facilities that simulate the operational environment is useful, data analysis can be done off-line, and, consequently, we have the luxury of dealing with measures based on derived indices such as average waveforms. Applications of this sort would include:

- Choosing among alternative hardware or software.

- Choosing among alternative procedures.

---

[1] Samaras, George M: Towards a Mathematical Formalism of Performance, Task Difficulty, and Activation. NASA CP 2504, 1988, pp. 43-55

- Assessing the fidelity of simulation.

- Use as a debriefing tool, to probe operators with additional questions, after-the-fact, about the times during a recorded scenario when the physiological signs suggested, for example, that the operator was stressed or distracted.

On-line, Real-time Applications. To the extent that physiological indices of performance can be extracted on one or a few trials (i.e., from single epoch recordings), and it is feasible to derive these indices in real-time in the operational setting, they would be useful in closed-loop man-machine systems. In general, this group of applications would involve the feedback of physiological information from the operator to the machine with which he is interacting, so that decision-making algorithms that reside there can modify the operator's task or displays accordingly. This group of applications is perhaps most demanding, because of the need for real-time turnaround of the measures of interest. Applications of this sort would include:

- Assessing the general state of the operator, to determine whether he is fit to be "in the loop" at all.

- Dynamically allocating tasks between the human operator and onboard AI, depending on workload.

- Checking whether the operator attended to events that the onboard AI flagged as significant, as well as detecting instances in which the operator realizes he made an error, so that he has an opportunity to correct himself.

Personnel Selection and Training. To the extent that physiological measures reflect cognitive processes for which there are significant individual differences, these measures may prove useful for selecting personnel and monitoring the progress of an individual's training. The challenge here is to define measures that are predictive of future performance. As with system design applications, we would frequently be able to process the recorded data off-line and deal with derived measures, without the constraints of real-time turnaround. Some applications of this type include:

- Staffing high workload tasks or environments with individuals who are well-suited to handle them.

- Channeling personnel into jobs that take advantage of their cognitive styles.

- Determining skills in an individual's training program that remain to be mastered by identifying the aspects of a task that cause high workload.

THE MOST PROMISING PHYSIOLOGICAL MEASURES

The research literature provides considerable evidence to suggest that a number of physiological measures will be useful for the applications mentioned above. It is beyond the scope of this paper to attempt a comprehensive review of this literature. However, in the present section a cursory overview is offered, to provide some indication of which indices of central and peripheral

sweats, and moves about in performing his duties. In addition, physiologically generated artifacts, most often from eye movements and blinks, or skeletal muscle activity, can likewise contaminate the recordings. While useful recordings of EEG have been reported in-flight (e.g., Refs. 10, 11), significant engineering advances are required in electrode application, signal processing, and artifact rejection before such recordings could be used routinely.

Event-related Potentials (ERPs). ERPs are also voltage fluctuations recorded from the scalp, but those which are time-locked to events, usually external stimuli. Transient ERPs are characterized by the amplitude, latency from stimulus onset, and scalp distribution of the various component peaks in the waveform. The stimulus-locked brain activity is typically examined after signal averaging over numerous presentations of the same event, although single trial analysis techniques are an active area of investigation. ERP recordings in operational settings are subject to the same technical constraints as those of ongoing EEG.

Various ERP components have been shown to vary reliably with cognitive processes (see review in Ref. 12), including selective attention (e.g., Ref. 13), expectancy (e.g., Ref. 14), discrimination processes (e.g., Ref. 15) and response preparation (e.g., Ref. 16). In contrast to the findings regarding ongoing EEG, there is a body of research that has shown very encouraging relationships between ERP indices and workload. This work, by Donchin, Wickens, and colleagues, is reviewed in the Munson,[2] et al. paper in the present Proceedings. There is evidence that ERPs may be used to reveal systematic cognitive effects in addition to those which are apparent from behavioral measures alone. For example, P300 latency has been shown to vary with only a subset of the manipulations that affect overt reaction time, suggesting that the timing of P300 indexes the completion of stimulus evaluation processes, independent of response selection processes (e.g., Ref. 17). In certain situations, P300 amplitude appears to be a reflection of subjective probability, whereas overt choice behavior may be influenced by additional variables, for example those which affect the willingness to take risks (e.g., Ref. 18).

Steady-state ERPs are recorded in response to a rapidly oscillating stimulus, usually a light or sound. They are usually quantified in terms of amplitude and phase delay at the frequency of stimulation, and can be calculated after only several seconds of stimulation. Steady-state ERPs elicited by rapid, periodic stimulation by a checkerboard have also been reported to reflect workload when the checkerboard was presented concurrently with task performance (e.g., Ref. 19). This result is surprising, given that steady-state responses had been previously thought to reflect strictly sensory processes. The effect needs to be further examined to rule out the possibility that peripheral changes in the visual system, such as accomodation, could be varying with task difficulty and thus mediating the changes in the steady-state response.

Electrooculography (EOG). EOG recordings are derived from electrodes on the face near the eyes and can be used to monitor eye movements, eye blinks, and,

[2]Munson, Robert C.; Horst, Richard L.; and Mahaffey, David L.: Primary TASK ERPs Related to Different Aspects of Information Processing. NASA CP 2504, pp. 163-178.

nervous system activity, of the many that can be recorded non-invasively from behaving humans, appear most promising for near-term application.

Although there is considerable overlap in the measures that appear useful for different kinds of applications, a distinction should be made between the use of physiological measures to indicate the basic fitness of the operator to perform his tasks and the use of physiological measures to infer cognitive status. The former applications entail the monitoring of vital signs to indicate relatively gross impairments in physical well-being — e.g., G-induced loss of consciousness or gray-out, exposure to CBR agents, motion sickness, heat stress, traumatic injury, heart attack. The latter applications entail the analysis of more subtle physiological changes related to task performance, so as to infer mental states such as high workload, fatigue, or inattention.

The following overview focuses on those measures with a demonstrated relationship to operationally defined manipulations of workload, stress, fatigue or boredom. While most of these relationships have been demonstrated in laboratory settings with non-real-time processing of the data, some have been recorded successfully in operational settings and all hold at least the promise of being feasible to derive in real-time. Typical quantitative measures that are derived from each physiological sign are presented, technical problems in recording these measures in operational settings are discussed, and examples of the evidence relating these measures to the psychological constructs of interest are mentioned. More extensive discussions of the prospects for using physiological measures in operational settings may be found in O'Donnell (Ref. 4) and Gomer (Ref. 5).

Electroencephalography (EEG). The EEG consists of voltage fluctuations recorded from two or more sites on the scalp, Ongoing EEG is usually quantified in terms of its frequency composition and amplitude asymmetries. Other measures, such as the coherence between the activity recorded at various pairs of scalp sites, also appear to be useful (e.g., Ref. 6).

Changes in the predominant frequencies in the EEG with levels of arousal and activation have been known for some time (e.g., Refs 7, 8). An alert person performing an engaging task shows predominantly low amplitude, fast frequency (beta) activity. An awake, but less alert, person shows an increased incidence of high amplitude, alpha (8-12 Hz) activity. With the onset of drowsiness, slower frequency theta (4-7 Hz) activity enters the spectrum and in the early stages of sleep, very high amplitude, slow (1-3 Hz), delta waves predominate. It is unlikely in operational settings that operators would lapse into deeper, so-called "paradoxical," stages of sleep. The generalized effect of stress, activation or arousal is, therefore, a shift towards the faster frequencies, often with an abrupt blocking of the alpha rhythm (e.g., Refs. 8, 9). Fatigue and boredom generally shift the spectrum in the other direction, towards the lower frequencies. Derived measures of ongoing EEG have not yet proven to be reliable indicators of workload.

Aside from the general problems of isolating the physiological recordings from environmental sources of electrical noise and deriving the measures of interest in near real-time, there are several technical problems in recording EEG and related measures in operational settings. Movement of the electrodes relative to the scalp causes severe electrical artifacts, and it is difficult to ensure firm contact in environments where the operator wears a helmet,

to a limited extent, direction of gaze and eye closure. The EOG reflects changes in the electric dipole formed between the cornea and the retina. While these potentials interfere with scalp-recorded electrophysiological measures such as EEG and ERPs, measures derived from the EOG itself have been shown to reflect operators' cognitive state.

Blink rate increases reflect the deterioration in attention and performance which occur over a prolonged task (e.g., Refs. 20, 21). Additionally, blink durations have been shown to increase with time on task (Ref. 22). Thus, increases in both blink rate and duration may indicate fatigue or lack of vigilance. As workload increases, blink rates decrease and the latency of the blink, after presentation of the stimulus of interest, increases (Ref. 23). Moreover, blinks during visual tasks were found to be of shorter duration than those in auditory tasks (Ref. 22). The pattern of these results are consistent with the notion that as visual information processing demands increase, eye blinks reflect the brain's attempt to take in more visual input.

Blinks are robust and easy to record, because they are of relatively high amplitude and predictable waveshape. Measures of blink frequency and latency should, therefore, be feasible even in somewhat noisy environments. Measures of blink duration will, of course, require relatively noise-free signals.

Eye Position and Pupil Dilation. Eye movements and fixations, and pupil dilation, are usually detected by photo-optical techniques and, therefore, are measures that can be gathered without sensors that touch the subject. Eye position is inferred from corneal reflectance and is usually quantified in terms of direction of gaze and dwell times as the eye scans the environment. Pupil size is measured in millimeters.

Dwell time on various displays on an instrument panel has been shown to vary systematically with workload (Ref. 24). Both tonic levels of pupil size over long durations of task performance and phasic responses elicited by task-relevant stimuli have been shown to be sensitive to cognitive variables. Tonic dilations seem to be a reliable index of activation and arousal (e.g., Ref. 9). In addition, consistent phasic increases in pupil dilation have been associated with increases in task difficulty and workload (e.g., Refs. 25, 26).

Because both these indices are dependent on maintaining a beam of light on the cornea, they are limited to environments, such as fixed-base simulators, in which there is minimal head movement by the operator. Eye trackers are becoming more sophisticated, but head movements beyond about one cubic foot take the eye out of range of the presently available photo-sensors. It is likewise difficult to maintain a fix on the eye in a high-vibration environment. Further confounds can be introduced by the fact that pupil size is responsive to non-specific factors such as ambient illumination, color, and depth of the visual field, which are difficult to control in operational settings.

Electrocardiography (ECG). ECGs are a widely used, easily recorded index of cardiovascular activity that is obtained from a two- or three-electrode array on the body. The ECG signal may be analyzed in terms of its basic timing (heart rate or period) or its morphology (e.g., amplitude of the T-wave). Derived measures from the ECG, given the detection of the R-wave as the basic datum, include first-order measures such as rate per unit time and change in

heart period across beats. Second-order analysis may include rate-of-change measures, maximum and minimum beat-to-beat periods within an epoch, and methods based on time-series analysis of the beat-to-beat intervals.

While heart rate has been shown to generally increase with stress (e.g., Ref. 27) and activation (see review in Ref. 9), the heart rate response to stimuli in a task environment is more often characterized by a complex pattern of deceleration and acceleration. The results of numerous (but not all) relevant studies are consistent with a hypothesis put forth by Lacey (Ref. 28), that heart rate deceleration reflects a receptivity to external stimulation whereas accelerations occur if the situation is found, after initial attention, to warrant an increase in energy release. Heart rate increases during periods of increased workload, for example during take-offs and landings, have been reported (e.g., Refs. 29, 30) but others have not found heart rate to be sensitive to the cognitive workload of simulated flight (Ref. 26).

More consistent relationships with workload have been reported for heart-rate variability. The general finding has been that, with increased attention and workload, heart-rate variability decreases (e.g., Refs. 31, 32). The most frequently used technique to reveal this workload effect has been a spectral analysis of the beat-to-beat time interval data with a focus on the power in the 0.1 Hz band (e.g., Ref. 33). Of particular interest has been the component of heart-rate variability related to respiratory sinus arrhythmia. because of the many influences on the beat-to-beat regularity of the heart, this one reflects mediation by the central nervous system. An approach to quantifying sinus arrhythmia, which makes fewer assumptions about the statistical properties (i.e., stationarity) of the data than those based on spectral analysis, is that of vagal tone. Porges[3] (see Ref. 34 and paper in this Proceedings) has developed a moving polynomial filter technique that removes the slowly shifting baseline from the inter-beat interval data over time in order to reveal the faster oscillations due to respiratory sinus arrhythmia. In the few instances in which this "vagal tone" measure has been compared to the measure based on power in the 0.1 Hz band, vagal tone has proven to be the more sensitive indicator of the experimental manipulations (Ref. 35).

Heart rate measures have been successfully recorded under extremely demanding conditions (e.g., Refs. 36, 37, 38).

Respiration. A number of techniques have been proposed for measurement of the basic respiratory signal. As a class, girth measurements of the thorax and/or the abdomen using mercury-in-silastic tubing strain gauges are simple, non-invasive, and reliable. If possible, both thoracic and abdominal components of the respiratory motion should be monitored, since it is possible to derive an adequate measure of respiratory volume from the combined signals. The principal measures are respiratory rate, average volume (if composite), and parameters related to the timing of inspiration, inspiratory pause, expiration, and expiratory pause. Tidal volume, the volume of air expired, can be sensed by thermistors mounted unobtrusively in an oxygen mask. Minute volume may vary independently of tidal volume and can be measured in the same way.

---

[3] Porges, Stephen W.: Vagal Tone as an Index of Mental State. NASA CP 2504, pp. 57-64

Respiration measures deserve more attention than they have received (e.g. Ref. 39) for detecting operator incapacity. There is also some indication that respiration becomes more shallow, regular and rapid with increased workload (Ref. 40).

Electromyography (EMG). EMG recordings from surface electrodes can be used to detect muscle tone or movement mediated by selected muscle groups, if they can be recorded without contamination by task-related movements. Several sites have been suggested as indicating overall tension levels, particularly forehead or masseter muscle placements. Since the signal is a complex, irregular one, the preferred strategy for determining general tension levels is to integrate the primary signal over a relatively short time constant, typically between 0.1 and 0.5 seconds, and to subsequently analyze only this average measure. The measures typically derived from the average muscle tension level are mean level, variance of the level, and minimum and maximum level for each epoch. If appropriate, further measures such as the number of increases above a criterion level can be obtained.

Muscle tension increases with arousal, stress and activation (e.g. Refs. 9, 41) and increased EMG activity is associated with the onset of fatigue. Several studies have reported relationships between increased EMG activity and increased workload or task difficulty (e.g., Refs. 42, 43), but it is as yet unclear as to how sensitive EMG is as an index of small changes in workload.

Other Measures of Interest. A number of other physiological measures deserve "honorable mention," either because they appear to be worthwhile indicants of cognitive status, but without the near-term prospects for application in the field, or because they appear to be related to cognition in only a general sense:

- Ongoing and stimulus-locked measures based on magnetoencephalography recordings are particularly promising because the sensor does not touch the subject's body and because inferences can often be made about the depth from which activity arises. Evoked magnetic fields have been correlated with attention and subjective probability in a paradigm similar to that used for ERP studies of P300 (Ref. 44). However, the sensors now in use must be supercooled with a large container of liquid helium and the subject must maintain a posture which keeps his orientation and distance from the sensor constant.

- Blood pressure and blood flow can provide useful information about cardiovascular status which, to some extent, complements that available from heart rate and heart rate variability. However, methods for recording these indices non-invasively have not yet reached the point that they would be useful in an electrically noisy, high vibration environment, or one in which the operator had to be free to move significantly.

- Advances are being made in the sensor technology for monitoring body temperature, with the development of miniaturized telemetry systems that can be swallowed as a "pill" and used to monitor core temperature as it passes through the gut, and with the development of improved skin temperature sensors. This technology promises to be of use in environments where heat stress is a threat, and phasic temperature changes have been related to mental workload (e.g., Ref. 45) as well as physical workload.

- Measures of skin resistance and skin conductance are relatively easy to record, and have some value for indicating phasic changes in arousal and stress, but they have yet to prove themselves as specific enough to be of utility for inferring cognitive states.

PROBLEM AREAS

There is no question that significant technical problems remain to be solved before physiological monitoring technology will come into widespread use in operational settings. But it is also apparent that recent technological developments offer new possibilities for solving many of these problems and that researchers, and funding agencies, are only now turning their attention towards these prospects. Some areas of concern are the following:

Instrumentation and Operator Acceptance. The operator's reluctance to be instrumented is an often-mentioned impediment to implementation of physiological monitoring in operational settings. Operators find conventional recording paraphernalia cumbersome and obtrusive. It is time-consuming to have electrodes pasted on and removed. They are also threatened by the possibility that in submitting to recordings, an unanticipated medical problem may be detected that could call into question their eligibility. When faced with the prospects of closed-loop decision-making, operators are reluctant to relinguish their control of a system to automated subsystems.

As recording instrumentation becomes more miniaturized, some of these objections will disappear. There are now several "pocket-size" amplifier/ recording systems available for ambulatory monitoring (e.g. the SSPIDR, see Banta's paper in this Proceedings). On-board storage of physiological data is now achieved with either cassette tape or solid-state memories. Optical disk media may soon provide still further storage capacity. Telemetry systems are likewise becoming smaller and more sophisticated. "Paste-less" electrodes have been a possibility for some time, but require further refinement. Integrating electrodes and amplifiers into helmets and uniforms remains a challenge, but is being addressed by several groups. The palatability of using physiological measures in closed-loop control systems will be increased by giving the operator the ability to override the decisions reached by the on-board decision-making algorithms, and by introducing this technology as an open-loop "aid" to the operator until the decision rules mature to the point that they warrant the operator's confidence. As for the objections which can't be addressed with instrumentation, one suspects that as the value of physiological measures becomes more apparent and the safety implications of not having them is more widely recognized, these problems will largely take care of themselves.

Safety issues. Any tethering of the pilot to recording equipment must be done in a way that does not distract or impede him from performing his duties. In some environments, such as fighter aircraft where the aircrew must be able to eject if necessary, this requirement dictates a telemetry system for transmitting the amplified physiological signals to on-board or remote processing equipment or an entirely portable physiological recording system that can be carried on the operator's person (e.g., Ref. 46). Furthermore, the recording equipment must be electrically integrated with the other equipment with which the operator interacts, so that there is no shock hazard when he touches the control stick or instrument panel.

Here again, advances in micro-electronics are allowing increased miniaturization, and thus portability, of amplifiers, storage media and telemetry systems. Amplifiers can be designed with fail-safe features to protect the subject against internal shorts in the circuitry, and the possibility of such failures can be minimized by "hardening" physiological recording equipment, using the same methods that are used for other on-board electronic instrumentation, for use in even high-vibration environments. If telemetry is used, it must be accomplished with a technique or in a frequency range that does not interfere with other on-board equipment. Ensuring against shock hazard involves issues of electrical grounding that can usually be readily solved with cooperation from system engineers.

Artifact Rejection and Compensation. There are two sets of issues regarding contamination of recordings by artifact — one involving electrical artifacts from the environment and the other involving physiological artifacts from the subject himself. Most operational settings are electrically noisy environments, so aside from the above safety issues, appropriate shielding and grounding must be implemented in order to get clean physiological recordings. Miniaturization of amplifier electronics and efforts to integrate this circuitry into helmets and suits, offers the prospects of placing the amplifier circuitry on or in close proximity to the electrodes, which should increase noise-immunity considerably. Such integration, which could include custom-fitting the electrode mounts for individual operators, will also minimize artifacts caused by even slight displacements of an electrode relative to the skin. Fortunately, some of the power supplies in fielded operational systems oscillate at frequencies considerably higher than the physiological signals of interest, so bandpass filters attentuate such noise sources more readily than the 60 Hz interference which can be a problem in the laboratory. Appropriate notch filters, akin to the 60 Hz filters used in many conventional amplifiers, can also be custom-designed for specific operational settings, as long as the frequencies being attenuated are sufficiently disparate from the physiological spectrum of interest.

Physiological artifacts from the operator himself can be more troublesome. As alluded to above, electrophysiological recordings of one physiological parameter can be contaminated by other physiological parameters with overlapping frequency components. For example, EEG and ERP recordings can be contaminated by eye blinks, heart beat, and muscle artifacts. Furthermore, excessive sweating can elicit skin potentials that interfere with the physiological measures of interest or can cause electrodes to be more easily dislodged. These problems dictate the need for innovative electrode designs, well-integrated into the operators clothing and other equipment, as well as the need for "intelligent" digital filtering algorithms (e.g., Ref. 47) to rid the recording of artifact.

Real-time Turnaround. As discussed in the "Areas of Application" section, many potential uses of physiological measures in operational settings do not require real-time turnaround of data analyses. In fact, most recordings to date in simulators or fielded systems have stored the amplified physiological signs on either analog or digital media for off-line analysis. Only recently have systems appeared with some on-board computing power (e.g. the SSPIDR), but even here the decision-making capability has thus far been limited to making intelligent decisions about when to store data into the limited-capacity memory for off-line analysis. The possibilities for real-time analysis of

physiological data, the use of derived measures for real-time decision-making, and the realization of closed-loop feedback based on the resultant decisions as an input to adaptive systems are areas that need to be pursued more aggressively. A reasonable way to proceed on this front would seem to be an initial focus on the development of pattern recognition algorithms for "single trial" extraction of useful indices from records of ongoing physiological activity, followed by non-real-time demonstrations of how these derived indices would be used for making useful decisions about operator status. Only then need there be an attempt to "speed-up" this process to real-time, perhaps by implementing the mature algorithms in special-purpose hardware.

Knowledge-based Interpretation of Physiological Measures. Whether or not real-time turnaround is required for certain applications in operational settings, there will certainly be the need for more automated means of interpreting physiological data than are presently available. Expert system techniques for encoding knowledge and applying decision rules offer possibilities as a framework for such automated interpretation, although it is not yet clear how complicated the decision-rules and contingencies will need to be. It is apparent, given the aforementioned cautions that have been raised about inferring mental states from physiological measures alone (Ref. 2), that it will be necessary to take into account simultaneously derived measures of operator behavior and system performance as a whole. Very little work has been done in modeling the integration of physiological, behavioral and system performance data. The paper by Samaras in the present Proceedings offers one possible framework for such an integration. Appropriate decision rules relating changes in physiological signs to mental states or predicted performance can be derived initially from the biomedical and psychophysiological literatures. However, refinements of these decision rules and proof-of-concept demonstrations will likely require the use of realistic scenarios in simulator environments.

SUMMARY OF AREAS FOR FURTHER DEVELOPMENT

The foregoing discussion has attempted to provide an overview of the state-of-the-art and the challenges that lie ahead "in the field," as physiological monitoring technology expands from the laboratory into operational settings. Although valid physiological measures have been recorded already in a number of demanding operational settings, including advanced cockpits, the methodologies for implementing such measures have been largely special-purpose and cumbersome. The successes to-date merely foreshadow the possibilities that exist, as conceptual and engineering advances continue. The following list summarizes a number of the areas that are fertile ground for further development:

- Advances in physiological sensor design and better ways of mounting electrodes in an operator's helmet, clothing, or other gear.

- Further miniaturization of amplifiers, digitizers, storage media, and telemetry equipment, along with design features to maximize noise-immunity and integration into the operator's physical environment.

- Digital filtering algorithms to minimize the contamination of recordings by artifacts, both those due to electrical sources in the environment and those due to physiological sources within the subject.

- Special purpose data analysis software or firmware that can process the recorded signals in near real-time.

- Modeling of mental states and task environments to allow physiological measures to be taken into account, along with behavioral, subjective, and system performance measures, in interpreting and predicting performance.

- Empirical work to develop decision algorithms for inferring the operational significance of operator physiological changes and for "closing the loop" between man and machine.

REFERENCES

1. Coles, M. G. H., Donchin, E. & Porges, S. W. (Eds.): Psychophysiology: Systems, Processes, and Applications. New York: Guilford, 1986.

2. Johnson, L. C.: Use of Physiological Measures to Monitor Operator State. In F. E. Gomer (Ed.), Biocybernetic Applications for Military Systems. Report No. MDC E2191, St. Louis: McDonnell Douglas Astronautics Company, 1980, pp. 101-131.

3. Zacharias, G. L.: Physiological Correlates of Mental Workload. Report No. 4308, Cambridge: Bolt, Beranek & Newman, Inc., 1980.

4. O'Donnell, R. D.: Contributions of Psychophysiological Techniques to Aircraft Design and Other Operational Problems. NATO AGARDograph No. 244, 1979.

5. Gomer, F. E. (Ed.): Biocybernetic Applications for Military Systems. Report No. MDC E2191, St. Louis: McDonnell Douglas Astronautics Company, 1980.

6. Gevins, A. S., Doyle, J. C., Cutillo, B. A., Schaffer, R. E., Tannehill, R. S., Ghannam, J. H., Gilcrease, V. A., & Yeager, C. L.: Electrical Potentials in Human Brain During Cognition: New Method Reveals Dynamic Patterns of Correlation. Science, vol. 213, 1981, pp. 918-922.

7. Lindsley, D. B.: Psychological Phenomena and the Electroencephalogram. Electroencephalography and Clinical Neurophysiology, vol. 4, 1952, pp. 443-456.

8. Malmo, R. B.: Activation: A Neurophysiological Dimension. Psychological Review, vol. 66, 1959, pp. 367-386.

9. Duffy, E.: Activation. In N. S. Greenfield & R. A. Steinbach (Eds.), Handbook of Psychophysiology. New York: Holt, Rinehart and Winston, 1972, pp. 577-622.

10. Sem-Jacobsen, C. W.: Electroencephalographic Study of Pilot Stresses in Flight. Aerospace Medicine, vol. 30, 1959, pp. 797-801.

11. Sem-Jacobsen, C. W., & Sem-Jacobsen, E. E.: Selection and Evaluation of Pilots for High Performance Aircraft and Spacecraft by Inflight EEG Study of Stress Tolerance. Aerospace Medicine, vol. 34, 1963, pp. 605-609.

12. Donchin, E., Ritter, W., & McCallum, W. C.: Cognitive Psychophysiology: The Endogenous Components of the ERP. In E. Callaway, P. Tueting, & S. H. Coslow (Eds.), Event-Related Brain Potentials in Man. New York: Academic Press, 1978, pp. 349-411.

13. Hillyard, S. A. & Hansen, J. C.: Attention: Electrophysiological Approaches. In M. G. H. Coles, E. Donchin, & S. W. Porges (Eds.), Psychophysiology: Systems, Processes and Applications. New York: Guilford Press, 1986, pp. 227-243.

14. Sutton, S. Braren, M., Zubin, J. & John, E. R.: Evoked Potential Correlates of Stimulus Uncertainty. Science, vol. 150, 1966, pp. 1187-1188.

15. Ritter, W., Simson, R., & Vaughan, H. G., Jr.: Association Cortex Potentials and Reaction Time in Auditory Discriminations. Electroencephalography and Clinical Neurophysiology, vol. 33, 1972, pp. 547-557.

16. Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C. & Winter, A. L.: Contingent Negative Variation: An Electrical Sign of Sensori-motor Association and Expectancy in the Human Brain. Nature, vol. 203, 1964, pp. 380-384.

17. McCarthy, G. & Donchin E.: A Metric for Thought: A Comparison of P300 Latency and Reaction Times. Science, vol. 211, 1981, pp. 77-79.

18. Karis, D., Chesney, G. L., & Donchin, E.: "...twas ten to one; And yet we ventured...": P300 and Decision Making. Psychophysiology, vol. 20, 1983, pp. 260-268.

19. Wilson, G. F. & O'Donnell, R. D.: Steady-State Evoked Responses: Correlations with Human Cognition. Psychophysiology, vol. 23, 1986, pp. 57-61.

20. Beideman, L. R. & Stern, J. A.: Aspects of the Eyeblink During Simulated Driving as a Function of Alcohol. Human Factors, vol. 19, 1977, pp. 73-77.

21. Bauer, L. O., Strock, B. D., Goldstein, R., Stern, J. A. & Walrath, L. C.: Auditory Discrimination and the Eyeblink. Psychophysiology, vol. 22, 1984, pp. 636-641.

22. Goldstein, R., Walrath, L. C., Stern, J. A., and Strock, B. D.: Blink Activity in a Discrimination Task as a Function of Stimulus Modality and Schedule Presentation. Psychophysiology, vol. 22, 1985, pp. 629-635.

23. Bauer, L. O., Goldstein, R., & Stern, J. A.: Effects of Information-Processing Demands on Physiological Response Patterns. Human Factors, vol. 29, 1987, pp. 213-234.

24. Harris, R. L., Sr., Tole, J. R., Stephens, A. T., & Ephrath, A. R.: Visual Scanning Behavior and Pilot Workload. Aviation, Space and Environmental Medicine, vol. 53, 1982, pp. 1067-1072.

25. Beatty, J.: Phasic Not Tonic Pupillary Responses Vary with Auditory Vigilance Performance. Psychophysiology, vol. 19., 1982, pp. 167-172.

26. Casali, J. G., & Wierwille, W. W.: A Comparison of Rating Scale, Secondary-Task, Physiological, and Primary-Task Workload Estimation Techniques in a Simulated Flight Task Emphasizing Communications Load. Human Factors, vol. 25, 1983, pp. 623-641.

27. Robinson, E. R. N.: Biotechnology Predictors of Physical Security Personnel Performance: I. A Review of the Stress Literature Related to Performance. Report No. NPRDC TN-83-9, Navy Personnel Research and Development Center, San Diego, Calif., 1983.

28. Lacey, J. I.: Somatic Response Patterning and Stress: Some Revisions of Activation Theory. In M. H. Appley & R. Trumbull (Eds.), Psychological Stress. New York: Appleton-Century-Crofts, 1967.

29. Roscoe, A. H.: Heart-Rate as an In-Flight Measure of Pilot Workload. In M. L. Frazier & R. B. Crombie, (Eds.), Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics. Edwards Air Force Base: AFTEC-TR-82-5, 1982, pp. 338-349.

30. Hart, S. G., Hauser, J. R., & Lester, P. T.: Inflight Evaluation of Four Measures of Pilot Workload. Proceedings of the Human Factors Society — 28th Annual Meeting, 1984, pp. 945-949.

31. Sayers, B. McA.: Physiological Consequences of Informational Load and Overload. In P.H. Venables & M.J. Christie (Eds.), Research in Psychophysiology. New York: Wiley, 1975, pp. 95 124.

32. Veldman, J. B. P., Mulder, L. J. M., Mulder, G., & van der Heide, D.: Attention, Effort and Sinus Arrhythmia: How Far Are We? In J. F. Orlebeke, G. Mulder, and L. J. van Doornen (Eds.), Psychophysiology of Cardiovascular Control. New York: Plenum Press, 1985, pp. 407-424.

33. Mulder, G.: Sinusarrhythmia and Mental Workload. In N. Moray (Ed.), Mental Workload: Its Theory and Measurement, New York: Plenum Press, 1979, pp. 327-344.

34. Porges, S. W.: Respiratory Sinus Arrhythmia: An Index of Vagal Tone. In J. F. Orlebeke, G. Mulder, and L. J. van Doornen (Eds.), Psychophysiology of Cardiovascular Control. New York: Plenum Press, 1985, pp. 437-450.

35. Hatch, J. P., Klatt, K., Porges, S. W., Schroeder-Jasheway, L. & Supik, J. D. The Relation Between Rhythmic Cardiovascular Variability and Reactivity to Orthostatic, Cognitive, and Cold Pressor Stress. Psychophysiology, vol. 23, 1986, pp. 48-56.

36. Burton, R. R.: Human Responses to Repeated High-G Simulated Aerial Combat Maneuvers. Aviation, Space, and Environmental Medicine, vol. 51, 1982, pp. 1185-1192.

37. Poppen, J. R. & Drinker, C. K.: Physiologic Effects and Possible Methods of Reducing Symptoms Produced by Rapid Changes in Speed and Direction of Airplane as Measured in Actual Flight. Applied Physiology, vol. 215, 1950.

38. Roman, J., Older, H., & Jones, W. L.: Flight Research Program: VII. Medical Monitoring of Navy Carrier Pilots in Combat. Aerospace Medicine, vol. 38, 1967, pp. 133-139.

39. Rugh, J. D., Wichman, H., & Faustman, W. O.: Inexpensive Technique to Record Respiration During Flight. Aviation, Space and Environmental Medicine, vol. 48, 1977, pp. 169-171.

40. Williges, R. C., & Wierwille, W. W.: Behavioral Measures of Aircrew Mental Workload. Human Factors, vol. 21, 1979, pp. 549-574.

41. Eason, R. G., Beardshall, A. & Jaffe, S.: Performance and Physiological Indicants of Activation in a Vigilance Situation. Perceptual and Motor Skills, vol. 20, 1965, pp. 3-13.

42. Corkindale, K. G., Cumming, F. G., & Hammerton-Fraser, A. M.: Physiological Assessment of Pilot Stress During Landing. In Measurement of Aircrew Performance, AGARD Conference Proceedings, CP#56, Brooks Air Force Base, Texas, 1983.

43. Jex, H. R. & Allen, R. W.: Research on a New Human Dynamic Response Test Battery; Part II. Psychophysiological Correlates. Proceedings of the Sixth Annual Conference on Manual Control, Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, 1970, pp. 743-777.

44. Okada, Y. C., Kaufman, L., & Williamson, S. J.: The Hippocampal Formation as a Source of the Slow Endogenous Potentials. Electroencephalography and Clincal Neurophysiology, vol. 55, 1983, pp. 417-426.

45. Hancock, P. A.: Task Categorization and the Limits of Human Performance in Extreme Heat. Aviation, Space, and Environmental Medicine, vol. 53, 1982, pp. 778-784.

46. Call, D. W., Kelly, D. M., & Robertson, D. G.: A Self-Contained, Man-Borne Biomedical Instrumentation System in the Flight Testing of Naval Weapons Systems. In M. L. Frazier & R. B. Crombie, (Eds.), Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics, Edwards Air Force Base: AFTEC-TR-82-5, 1982, pp. 318-321.

47. Widrow, B., Glover, J. R., McCool, J. M., Kaunitz, J., Williams, C. S., Hearn, R. H., Zeidler, J. R., Dong, E., & Goodlin, R. C.: Adaptive Noise Cancelling: Principles and Applications. Proceedings of the IEEE, vol. 63, 1975, pp. 1692-1716.

N88-23372

S2-54
140790
13P.

Toward a Mathematical Formalism of Performance, Task Difficulty, and Activation

George M. Samaras
GMS Engineering Corporation
Columbia, Maryland

G1063871

CN+ G6

INTRODUCTION

Both people and their environments are **reciprocal** *determinants of each other.*
A. Bandura (Social Learning Theory, 1977)

The continually evolving sophistication and complexity of military and civilian technology is increasing the burden on human operators in man-machine systems. Whether a weapons platform or space vehicle, a power plant or factory control station, or even an aid for the handicapped, the informational and operational demands will ultimately exceed human capabilities, unless the man can be relieved by the machine. Dynamic task partitioning, shifting and sharing tasks between human and machine in real time is theoretically feasible. However, it is currently impossible to implement, since the man-machine interface lacks **reciprocal** status assessment capability. This lack of reciprocity is a key indicator of the low level of man-machine integration and results in the realization that the interface is a weak link, which can directly degrade mission success and jeopardize system survival.

In order to achieve reciprocal status assessment, it is necessary to provide means for the machine to monitor the human, while continuing to improve the means by which the human monitors the machine. Assessment of human functional status should include both physical and mental-state estimation, which may be approached by physiological and behavioral monitoring. While this is presumed necessary, is it sufficient? Functional status, of human or machine, is only operationally relevant in the context of predicting performance - for our ultimate end point is to maximize system performance, while conserving valuable resources (men, machines, and information). Therefore, functional status is an input for predicting performance, survival and mission success.

Workload is frequently offered as a means of evaluating system design and predicting system performance, survival, and mission success. But the term "workload" has numerous connotations (ref. 1) and, rather than referring to a well-defined, unique, and generally agreed upon phenomenon, it serves as a convenient label for a number of events, ideas, states, dimensions and other constructs that are ill-defined and difficult to measure (ref. 2). Sheridan and Stassen (ref. 1) have illustrated six alternative definitions (D1 - D6) and four corresponding measurements (M1 - M4) of "workload" in a control paradigm (see Figure 1). Clearly, only one (or none) of these definitions is scientifically permissible. Part of this dilemma may be circumvented by operationally segmenting "workload" into physical (D4) and mental components, reducing the candidate set of definitions for "mental workload" to five possibilities. Performance (D6) is not "workload", further reducing the candidate set to four. An attempt could be made to further segment "mental workload" into objective, operator-independent (D1 & D2) and subjective, operator-dependent (D3 & D5) components. However, D1 and D2 are not independent of the person performing the task; even the most well-intentioned individuals covertly corrupt (interpret) their assigned tasks and performance criteria, based on their perception of their organization's "reward structure" - which is, unfortunately, temporally unstable, because organizations are usually diachronically and synchronically inconsistent.

Given the definitional problems of "workload", it is theoretically and practically not useful if the objective is to realize an *engineering* solution for the problem of predicting man-machine system performance, survival and mission success. There may be numerous alternative approaches for solving this problem. One potentially useful path is to invoke the relatively old Yerkes-Dodson postulate (which purports to relate performance as a function of task difficulty and activation (**refs.** 3 and 4)) (see Figure 2) and the relatively new psycho-technology of cognitive behaviorism (Organizational Behavior Management, which purports to be a systematic, structured approach to human performance problem-solving (e.g. **ref.** 5)). Let us assert that performance is what is important, in the practical world of military and civilian operations, and that if performance is maximized, while minimizing the loss of valuable resources, the same endpoint is obtained as if it were practical to define, measure and control "workload". This "end run" around "workload" requires definition of performance, task difficulty, and activation in a manner useful to the system designer - an engineer normally lacking extensive training in physiology and psychology.

PERFORMANCE, TASK DIFFICULTY AND ACTIVATION

The rudiments of a mathematical formalism for integrating system performance, task difficulty, and physiological activation are offered here with the explicit understanding that it is unnecessary for this formalism to be correct or true - but that it is essential for the formalism to be useful! The implication here is that a *technology* is under development, which is to be evaluated by its effectiveness, as opposed to a *science*, which must be evaluated by the correctness of its theories. The purpose of this mathematical formalism, which employs existing mathematical tools that are well known to engineers, is to provide a framework for developing a structured, systematic approach **for:**

a) communicating physiological and psychological requirements, in a qualitative and quantitative manner, to the system design engineer, and
b) simplifying the problem of instructing a machine in the measurement and utilization of performance.

Basic Definitions

Define a mission ($\underline{M}$) as an ordered set of m explicit goals (G_i), such that:

$$\underline{M} = \{\underline{G}_1, \underline{G}_2, \underline{G}_3, \ldots, \underline{G}_i, \ldots \underline{G}_m\} \qquad \text{[Eqn. 1.01]}$$

A mission segment, a commonly used term, can then be viewed as a subset of these goals. In this formalism, a mission cannot exist unless one or more explicitly defined goals exist and it follows that mission performance cannot exist without goal performance. The term *explicit* is used in the same fashion as Farina & Wheaton (ref. 6); explicit means a goal was presented to, at least, the operator and one independent observer (not necessarily human) and that some objective procedure exists, allowing the observer to verify whether or not a goal has been achieved. A specific goal ($\underline{G}_i$) is then defined to be a function of a specific task ($\underline{T}_i$) and a task-specific criterion (C_i).

A task will be viewed as a position vector in some N-dimensional, time independent, state space ($\mathbf{D}^N$), such that the task describes the difference ($\Delta\underline{S}^g$) in **position** between the goal state ($\underline{S}^g$) and the origin state ($\underline{S}^o$) in the, usually local, environment.

$$\underline{T} = \Delta\underline{S}^g = \underline{S}^g - \underline{S}^o \qquad \text{[Eqn. 1.02]}$$

A task is thus defined as a criterion-independent **vector** variable that is solely a function of the component dimensions of $\mathbf{D}^N$. In order to simplify this exposition, it is explicitly assumed that $\underline{S}^g$ is an idealized point, rather than a volume, in task space. This allows consideration of performance only relative to a criterion of time. Considerations of performance relative to

variations in the task (the goal state as a volume, instead of a point) are also appropriate, but only **make** this exposition more complex - without contributing additional conceptual information.

A criterion (C_i) is defined as a time-dependent scalar variable that is independent of $\mathrm{Đ}^N$ and will be viewed as the time lapse (Δt^g) that is required for translation from the origin state ($\underline{S}^o$) to the goal state ($\underline{S}^g$), in order to complete the task and attain the goal.

$$C = \Delta t^g = t^g - t^o \qquad \text{[Eqn. 1.03]}$$

A goal is then defined as the algebraic ratio of a task and a task-specific criterion.

$$\underline{G}_i = \underline{T}_i/C_i = (\Delta\underline{S}^g)_i/(\Delta t^g)_i \qquad \text{[Eqn. 1.04]}$$

The analogous construct in classical physics is velocity, which is the time rate of change of position in space; it is the ratio of a position vector and time. In this formalism, goals will be conceptualized analogous to mean velocities, tasks analogous to displacements and criteria as time lapses (until the goal state is generalized from a point to a volume).

Conceptualizing a task as a displacement in the environmental state - from origin state to goal state, it is further recognized that:

a) a task is a change in state which is the consequence of time-dependent behaviors (overt or covert and voluntary or involuntary), just as a "physical" displacement is a consequence of (time-dependent) velocities;
b) a task may be characterized according to its difficulty, just as a "physical" displacement may be characterized according path-dependent dissipative effects; and
c) a task requires physical and/or mental energy release, just as a "physical" displacement requires work.

Equation 1.04 describes a goal as a mean velocity across a geometrically minimum (presumed optimal) path from origin state to goal state. Given that the integral state change is the consequence of time-dependent behavior(s), the instantaneous temporal rate of change in state, at any instant, is construed as the vector variable behavior ($\underline{B}$). Thus,

$$\underline{B} = d\underline{S}/dt \qquad \text{but} \qquad \underline{G}_i = (\Delta\underline{S}^g)_i/(\Delta t^g)_i$$

Decomposing the resultant vector into orthogonal components, with one component ($\underline{r}^g$) having the same direction as the goal vector ($\underline{G}_i$), yields a goal-directed vector component ($\underline{B}^g$) that will be termed *purposive* behavior.

$$\underline{B}^g = d\underline{r}^g/dt \qquad \text{[Eqn. 1.05]}$$

A benefit of this approach is that, while an "instantaneous goal" can have no meaning, progress (both direction and magnitude) toward or away from a goal may be determined at any point in time. This lays the foundation for predicting whether or not the goal state will be achieved within the time criterion. Furthermore, it begins to permit determination of whether the operator is "leading" or "lagging", so that "leveling" via dynamic task partitioning can be implemented:

a) if the operator is "lagging" the goal trajectory, then assistance in various forms can be provided to "lighten the load"; or
b) if the operator is "leading" the goal trajectory, then slack time will result which may be used for lower priority goals, including preventing boredom or decrements in vigilance.

At this juncture, a few clarifications are required. First, what are the dimensions of the task space and is it necessary to identify all of the task dimensions for any given task?

Let us assert that only those dimensions containing critical features of the task need to be identified; other dimensions, where variation on these dimensions does not lead to significant redefinition of the task, can (in the first approximation) be ignored. This is not an example of logical positivism, but merely a standard engineering ploy to capture the important aspects of a process/problem without unnecessarily expending resources on higher order effects. Thus, the definition of the task determines the dimensions of the task space. Second, doesn't this formalism fail in the case of a "tracking" task (e.g. just maintain a constant altitude), where the goal state and the origin state are the same? Doesn't this imply that the goal does not exist, since the task is zero ($\Delta\underline{S}^g = \underline{S}^g - \underline{S}^o$)? No, quite the contrary. The goal does **exist**, and the goal is to have a zero change in altitude (tasks have direction and magnitude) in the specified time period.

Performance

Performance is defined as a scalar variable whose functional form will depend on assessment of the values assigned to various alternative outcomes. This is a classical problem of operations research and can be approached by standard decision theory and utility theory techniques, with the aid of probabilistic risk assessment. While the details are beyond the limited scope of this exposition, let us assume that the decision-maker's "utility" function (performance versus outcome) has been determined, either by direct measurement or by any one of a number of standard indirect methods, and has the following form:

$$P = f\,[\underline{\mathbf{G}},\underline{\mathbf{B}}(t)] = e^{-[x/\lambda]^2}$$

where:

$$x = \left\{(^1/_{\Delta t}) \int_{t^o}^{t} \underline{\mathbf{B}}(t)\,dt\right\} - \underline{\mathbf{G}}$$

and λ is some shape factor, $\underline{\mathbf{B}}$ is the measured behavior, and $\underline{\mathbf{G}}$ is the goal. This functional form is no more than that of a normal distribution and was selected somewhat arbitrarily. It is by no means the *only* form nor is it the *correct* form of the performance function; the correct form can only be that form chosen (directly or indirectly) by the decision-maker responsible for setting the goal and defining performance. It does, however, have some interesting properties:

a) it is a continuous function with range $0 \rightarrow 1$ and infinite domain (all possible outcomes);

b) it is symmetrical about $x = 0$, the implication being that reaching the goal state too early (*wasting fuel*) is just as bad as arriving too late (*missing the rendezvous*); and

c) when the value of $x = 0$, performance is 1.0 and as $|x|$ increases in magnitude, performance decreases towards zero.

The specific functional form of performance has not been defined, since it may vary with each goal and each decision-maker. However, a mathematical basis for completely determining its functional form, independent of the operator and using standard tools **has been defined**. While, at first, this appears to place an unreasonable burden on the organization defining the mission, this is not true. Both military and civilian organizations are constantly striving to structure operations and define objectives. For any specific man-machine system (SC/AT* **heli**copter, sonar/radar system, nuclear power plant, etc.) the number and diversity of tasks and goals are finite and considerably constrained. Therefore, not only is the problem tractable, but clear definitions of tasks, time criteria and performance measures are an integral and necessary part of effective and efficient communication of the mission objectives to the human operator.

* Scout/Attack (SC/AT)

Difficulty

Let us view task or goal difficulty as a construct that impedes goal attainment. A number of investigators have proposed "dimensions" for characterizing human operator tasks. One example is that of Farina & Wheaton as described by Fleishman & Quaintance (ref. 7) and contains 21 "dimensions" and associated measuring scales, with range $1 \rightarrow 7$. In this formalism, some of these dimensions will be used to develop a scalar coefficient termed task or goal *difficulty*, in keeping with the Yerkes-Dodson principle requiring performance to be a function of task difficulty and activation. Fleishman & Quaintance (ref. 7) cite examples in which polynomial constructs using various of these dimensions have been correlated with performance - a result expected based on the Yerkes-Dodson principle. Table 1 enumerates the original 21 candidate dimensions and identifies four which do not appear independent (items [4], [5], [13], and [20]). Since orthogonality is essential, only the remaining 17 appear acceptable. Furthermore, consistent with this formalism, candidate dimension [2] is recognized as time-dependent and thus permissible for constructing goal difficulty, but not task difficulty. Task difficulty is then defined using a weighted combination of the 16 remaining dimensions; goal difficulty (ξ) is defined when the 17^{th} criterion-based dimension, [2], is included in the combination. There are two classical forms for constructing such a combination, a weighted sum or a weighted product:

$$\xi = \sum \beta_k X_k \qquad \text{or} \qquad \xi = \prod \beta_k X_k \qquad \text{[Eqn. 1.06]}$$

where k = dimensional identifier ($1 \rightarrow 17$), β_k = regression coefficients from a population of operators, and X_k = an individual operator's rating ($1 \rightarrow 7$ using the existing rating scales or 0, if the dimension is not relevant), so that individual differences can be accommodated. Discriminating between these two functional forms, or some intermediate form, is a classical problem; consider, for example, the well-known Valency-Instrumentality-Expectancy (VIE) theory (refs. 8 and 9), where both forms often correlate well with the intervening variable. Selection of the preferred functional form of ξ must await empirical investigation.

Once again, as in the case of performance, this formalism does not provide a simple answer for determining task or goal difficulty. Difficulty is expected to vary with the individual operator and the specific goal. However, the formalism does provide a structured, systematic means of determining difficulty that may allow psychologists to communicate to engineers **quantitative** information that can be employed in the system design, development, and implementation process.

Activation

Every *living* organism exists in a state of dynamic quasi-equilibrium and may be viewed as an energy transducer - obtaining, storing, and releasing energy in different forms. This release of stored energy results in the production of work and heat which may (directly or indirectly) be detected in the form of behaviors (overt or covert) having magnitude (intensity) and direction (goal-directed or otherwise). The concepts of arousal (phasic) and activation (tonic) have their origins at least as early as the beginning of this century, when attempts were made to relate variations in behavioral intensity and performance to variations in psychophysiological activity (ref. 10). This work suggested that behavior could be regarded as varying along a continuum of intensity, from deep sleep to extreme excitement, and attempts were made to specify the physiological changes taking place at crucial points on this continuum - which became known as the level of activation or arousal (refs. 11, 12, and 13).

If the premise that behavior, as defined, requires the release of energy, the existence of a continuum can be logically deduced. At one extreme, a *living* organism must expend some minimal energy to sustain fundamental life processes. At the other extreme, there must be some maximum release rate beyond which the organism will be destroyed due, if nothing else,

to its inability to shed heat rapidly enough to prevent thermal denaturation of its constituent macromolecules. Between these limits, a variety of release rates are expected as the organism attempts to cope, as best it can, with the vagaries of its environment.

Merely employing the total energy release rate as an index of activation, while attractive in its simplicity, ignores an intrinsic property of the organism - the homeostatic tendency that operates over a reasonably wide dynamic range, that tends to maintain the organism in a state of dynamic quasi-equilibrium, and that arises because the organism is, as Sherrington* **indicated,** integrated. In the absence of changing external (environmental) and internal (needs, drives) forces, the organism will generally waver about the same release rate. Conversely, in the presence of changing external or internal forces, the level of energy release changes until the forces acting on the organism abate.

This wavering, in the "relaxed" state, is probably due to the looseness (wide deadband) of the organism's internal feedback control systems; candidate physiological measures of arousal or activation - taken while subjects were simply doing nothing in a relaxed state - were found to have fairly low positive correlations. However, when the system is driven (a standard engineering ploy in systems analysis) so that arousal is presumably induced, the candidate measures change in the expected direction. An example is Berlyne's meta-analysis (ref. 14) of several studies on mental effort; average EEG frequency, muscle tension, heart rate and skin conductance increased with purported increases in mental effort. Furthermore, Eason & Dudley (ref. 15) measured EEG evoked potentials, heart rate, skin resistance and muscle tension and reported that, with increasing task difficulty (they presumed this to be more arousing), the greater the change and all measures acted together.

The activation phenomenon, however, is not simple. Physiological indices that *hypothetically* measure arousal or activation actually move in different directions for different tasks. During tasks that require intake of information, Lacey (ref. 16) has shown that heart rate decreases while skin conductance increases. Alternatively, with tasks requiring internal processing or thinking, the reverse has been reported. What appears implicit from these findings is that careful consideration of the underlying energetics, from the organism's point of view, is imperative. Simply monitoring a physiological or behavioral parameter, without consideration of the specific operational circumstances, should not be expected to yield useful information.

In this formalism, activation energy (A) level is defined as a scalar variable, the resultant level of energy release derived from a weighted set of physiological (Φ) and behavioral (Ψ) measures. One possible form is:

$$A = \sum \gamma_j Y_j \qquad \text{[Eqn. 1.07]}$$

where j = the Φ or Ψ measure specifier, γ_j = bipolar weighting factors, and Y_j = the preprocessed Φ or Ψ data. It must be obtained while the human is being driven, not by operationally irrelevant secondary tasks, but during the normal control cycle of a dynamic task partitioner that is shifting and sharing mission relevant tasks between man and machine. Furthermore, the sign of the bipolar weighting factors must be determined based on rules that integrate the specific physiological and behavioral measures with the specific task(s) or, more realistically, task categories. Such rules, except for very simple cases, are currently undetermined. However, it is not unreasonable to expect that, in the presence of well-defined tasks and an appropriate set of physiological/behavioral measures, such rules can be developed from physiological principles and energetic considerations. Whether or not activation and difficulty, as defined here, will provide a robust estimate of performance can only be determined empirically.

*Sherrington, C.S.: Integrative action of the nervous system. New Haven: Yale University Press, 1906.

Derivative Quantities

We have defined quantities analogous (in physics) to time interval (criterion), displacement (task), mean velocity (goal), and instantaneous velocity (behavior or behavioral component). Much of classical physics deals with a large number of physical quantities that can be expressed in terms of a very small number of arbitrarily defined "articles of faith"; the fundament of physics is the existence of mass (m), length (l), and time (t). Each of these is arbitrarily defined and a standard quantity of each, agreed upon by most scientists, is maintained for reference in Paris and Gaithersburg. With these standards and the *principle of concatenation* we are able to determine other masses, lengths, and times, as well as derivative quantities. Examples of some derivative quantities, in terms of m,l,t are: area (l^2), volume (l^3), velocity (lt^{-1}), acceleration (lt^{-2}), density (ml^{-3}), momentum (mlt^{-1}), force (mlt^{-2}), energy (ml^2t^{-2}), frequency (t^{-1}), angular momentum (ml^2t^{-1}), and pressure ($ml^{-1}t^{-2}$). Even electric charge (q) was measured in terms of these basic and arbitrary quantities - through the ingenious Millikan oil drop experiment.

What this implies is that, no matter how complex the physical phenomenon, measurements can only be made in the very small number of arbitrarily defined dimensions that underlie the nomological network of classical physics. Analogously, this mathematical formalism requires a similar set of fundamental dimensions. Time and length have already been proposed as the underlying dimensions for criterion, task, goal, and behavior. However, without a hypothetical construct analogous to physical mass, more complex derivative quantities are prevented.

In this formalism, motivation ($\underline{\mathbf{M}}$) will is defined as a vector variable and an acceleration analogue, in that changes in behavior can be construed to be the consequence of motivation. In physics, the existence of acceleration requires the existence of force(s) - actually a net force. Invoking the principle of continuity of cognitive behaviorism, external (environmental) forces will be recognized as creating internal (need or drive) forces ($\underline{\mathbf{N}}$) which result in motivation. Can motivation or needs be directly measured? No, but then forces and acceleration cannot be directly measured; only mass, length and time can be measured!

Theorists in motivational psychology have postulated that performance is a function of the product of ability and motivation (ref. 8). This is consistent with the proposed formalism, if ability is considered as a mass analogue, since a need or drive would create motivation which would create a change in behavior leading to a displacement in task space. It would then become possible to conclude that, for a given ability, the greater the need, the greater the resultant motivation. Conversely, for an **observed** motivation, the less the ability, the greater the need. This latter statement initially appears counter-intuitive. However, in this formalism ability (α) is defined as a scalar variable that includes not only genetically determined (physical and mental) aptitude as well as experience and training, but also self-concept (a variable traditionally included in motivation). Therefore, if one's expectancy is that one cannot execute a task, then (in order to obtain the same level of motivation) it will require a greater need/drive than if one's expectancy was that one was quite proficient (and that the requisite behavior would lead to accomplishing the task, that one wanted to emit the requisite behavior, and that one wanted the reward - in other words, VIE theory).

In this formalism, it is postulated that the vector variables $\underline{\mathbf{N}}_\ell$ and $\underline{\mathbf{M}}_\ell$ are functionally related by the scalar variable α, such that:

$$\underline{\mathbf{M}}_\ell = \alpha \, \underline{\mathbf{N}}_\ell \qquad \text{[Eqn. 1.08]}$$

It is presumed that over reasonable time intervals, the magnitude of α should remain relatively stable (time independent). However, in the presence of fatigue, boredom, stress, or injury,

apparent ability decreases. Therefore, it may be useful to define α - over reasonable time intervals - as the product of two variables, α_i and α_e, such that:

$$\alpha = \alpha_i \, \alpha_e \qquad \text{[Eqn. 1.09]}$$

where α_i has a relatively stable (*intrinsic*) value for a given individual and α_e is unstable (*extrinsic*) depending on fatigue, etc.

How can α_i be measured? In classical physics, the mass of an unknown object is found by comparison of its behavior to the behavior of an arbitrarily defined reference mass (the principle of concatenation). By analogy, it is therefore possible to define α_i in terms of some *arbitrary* reference ability. Of course, this raises the problem of how to apply a standard "force" in order to permit the determination; but this is no greater a problem than that found in classical physics. It can be solved by ingenuity, just like Millikan and his oil drop experiment! One potentially useful approach may be the Ability Rating Scale approach cited by Fleishman & Quaintance (ref. 7). Furthermore, one approach for determining α_e, in real time, may be a variant of Schmidtke's theory of destabilization classification in which fatigue is staged based on changes in the mean and variance of performance (ref. 17).

As originally stated, only the **rudiments** of a formalism are offered here. This mathematical structure (and associated measurement procedures) is far from complete. But there may be considerable power in this approach as indicated in the following simple example. Work is a path dependent function. Transition from an origin state to a goal state can be characterized by a minimum energy trajectory - this "optimum" path having been defined by the goal. Based on this, the goal-directed work requirement ($\underline{W}^g$) can be computed as:

$$\underline{W}^g = \xi \int_{S^o}^{S^g} \underline{N}^g d\underline{S} \equiv \xi \int_{S^o}^{S^g} \alpha^{-1} \underline{M}^g \, d\underline{S} \equiv \xi \int_{S^o}^{S^g} \alpha^{-1} (d^2 \underline{r}^g / dt^2) d\underline{S} \qquad \text{[Eqn. 1.10]}$$

which is expressed solely in terms of task, time, ability, and the subjective difficulty scale factor (ξ). This is not the actual work expended to attain the goal, as work will vary depending on the specific path taken; instead it may be viewed as the minimum increment (decrement) in work resulting from including (deleting) this particular goal in (from) the mission. $\underline{W}^g$ is an important quantity for any decision algorithm attempting to dynamically partition predetermined tasks between man and machine or to modify tasks in "mid-flight".

CONCLUSIONS

The rudiments of a mathematical formalism for handling operational, physiological, and psychological concepts have been developed for use by the man-machine system design engineer. The mathematical formalism provides a framework for developing a structured, systematic approach to the interface design problem, using existing mathematical tools, and simplifying the problem of "telling" a machine how to measure and use performance. If this formalism proves useful, the wealth of human knowledge in mathematics and physics can be transported, at very little cost, to solving problems in this area.

Figure 3 presents a diagrammatic means of envisioning how an "expert" metacontrol unit might be implemented within a man-machine system (ref. 17). Physical data from the machine (via its data bus) are acquired and preprocessed; physiological and behavioral data from the operator (via appropriate sensors) are acquired and preprocessed. These dynamic data are periodically introduced into the knowledge base, which also contains **machine** attributes (from the machine developer), **human** attributes (from biomedical/training personnel), **mission** attributes (from the mission planners), **operator** attributes (from simulator training), the rules of a complete "mathematical formalism", the rules of the OBM interventions, and other relevant deterministic and stochastic information. An inference engine utilizes this knowledge base to decide how to

partition tasks between man and machine to maintain maximum *system* performance with the minimum cost in valuable resources.

ACKNOWLEDGEMENT

This work was supported in part by U. S. Army Medical Research Acquisition Activity Contract No. DAMD17-86-C-6027 entitled "Conceptual Design of a Biocybernetic Link for Workload Leveling via Dynamic Task Partitioning", a collaborative effort of GMS Engineering Corporation and ARD Corporation, Columbia, MD. The opinions expressed here are those of the author and are not to be construed as an official position of the U.S. Army or N.A.S.A., unless so designated by other authorized documents.

REFERENCES

1 Sheridan, T.B. and Stassen, H.G.: Definitions, Models, and Measures of Human Workload, in **Mental Workload, Its Theory and Measurement**, N. Moray, (Ed.), Plenum Press, New York, 1979

2 Hart, S.G., Childress, M.E., and Hauser, J.R.: Individual Definitions of the Term "Workload", **1982 Psychology in the DOD Symposium**, 1982

3 Yerkes, R.M. and Dodson, J.D.: The relation of strength of stimulus to rapidity of habit formation, **J. Comp. Neurol. Psychol.**, 18:459-482, 1908

4 Martindale, C.: **Cognition and Consciousness**, Dorsey Press, Illinois, 1981

5 Luthans, F. and Kreitner, R.: **Organizational Behavior Modification and Beyond**, Scott, Foresman and Co., Glenview, Illinois, 1985

6 Farina, A.J. and Wheaton, G.R.: Development of a taxonomy of human performance: The task characteristics approach to performance prediction, **JSAS Catalog of Selected Documents in Psychology**, 3:26-27 (Ms. No. 323), 1973 (cited in Fleishman & Quaintance, 1984)

7 Fleishman, E.A. and Quaintance, M.K.: **Taxonomies of Human Performance - The Description of Human Tasks**, Academic Press, New York, 1984

8 Lawler, E.E.: **Motivation in Work Organizations**, Brooks/Cole, Monterey, California, 1973

9 Pinder, C.C.: **Work Motivation - Theories, Issues, and Applications**, Scott, Foresman and Co., Glenview, Illinois, 1984

10 Duffy, E.: The relationship between muscular tension and quality of performance, **Am. J. Psychol.**, 44:535-546, 1932

11 Duffy, E.: **Activation and Behavior**, Wiley, New York, 1962

12 Lindsley, D.B.: Emotion, in **Handbook of Experimental Psychology**, S.S. Stevens, (Ed.), Wiley, New York, 1951

13 Malmo, R.B.: Activation: a neuropsychological dimension, **Psychol. Rev.**, 66:367-386, 1959

14 Berlyne, D.E.: **Structure and Direction in Thinking**, Wiley, New York, 1965

15 Eason, R.G. and Dudley, L.M.: Physiological and Behavioral Indicants of Activation, **Psychophysiology**, 7:223-232, 1971

16 Lacey, J.I., Somatic response patterning and stress: Some revisions of activation theory, in **Psychological Stress**, M.H. Appley & R. Trumbull, (Eds.), Appleton-Century-Crofts, New York, 1967

17 Samaras, G.M. and Horst, R.L.: Conceptual Design of a Biocybernetic Link for "Workload" Leveling via Dynamic Task Partitioning, Final Report, USAMRDC Contract No. DAMD17-86-C-6027, July 1986

Table 1: Task Characteristics
(adapted from Fleishman & Quaintance, (ref. 7), pgs 474-494)

[1] **number of output units** - an output unit is what is produced by the task

[2] **duration for which an output must be maintained** - in our terminology this is a criterion which, together with a task, defines a goal

[3] **number of elements per output unit** - elements are the parts or components which comprise the output unit

[4] **workload** - defined as a function of the number of output units [1] to be produced relative to the time [2] allowed for their production or the length of time for which an output must be maintained

[5] **difficulty of goal attainment** - defined as a function of [3] and [4] and thus not an independent dimension

[6] **precision of responses** - the degree to which fine or exacting responses are required

[7] **response rate** - the frequency with which responses must be made

[8] **simultaneity of responses** - the number of effectors (e.g. hand, foot, arm, voice) used for responding in order to produce an output unit (mental activities are not included here, but are in item [21])

[9] **degree of muscular effort involved**

[10] **number of procedural steps** - the number of responses needed to produce one output unit

[11] **dependency of procedural steps** - the degree of sequencing or linkage of procedural steps required

[12] **adherence to procedures** - the degree of criticality of following a prescribed sequence and stated procedures

[13] **procedural complexity** - defined as a function of [10] and [11]

[14] **variability of stimulus location** - the predictability of the physical location of the stimulus or stimulus complex

[15] **stimulus or stimulus-complex duration** - the fraction of time that the stimulus or stimulus-complex is available

[16] **regularity of stimulus occurrence** - the duration of inter-stimulus intervals (constant presence is considered equivalent to regular interval) and is a measure of the randomness of stimulus presentation

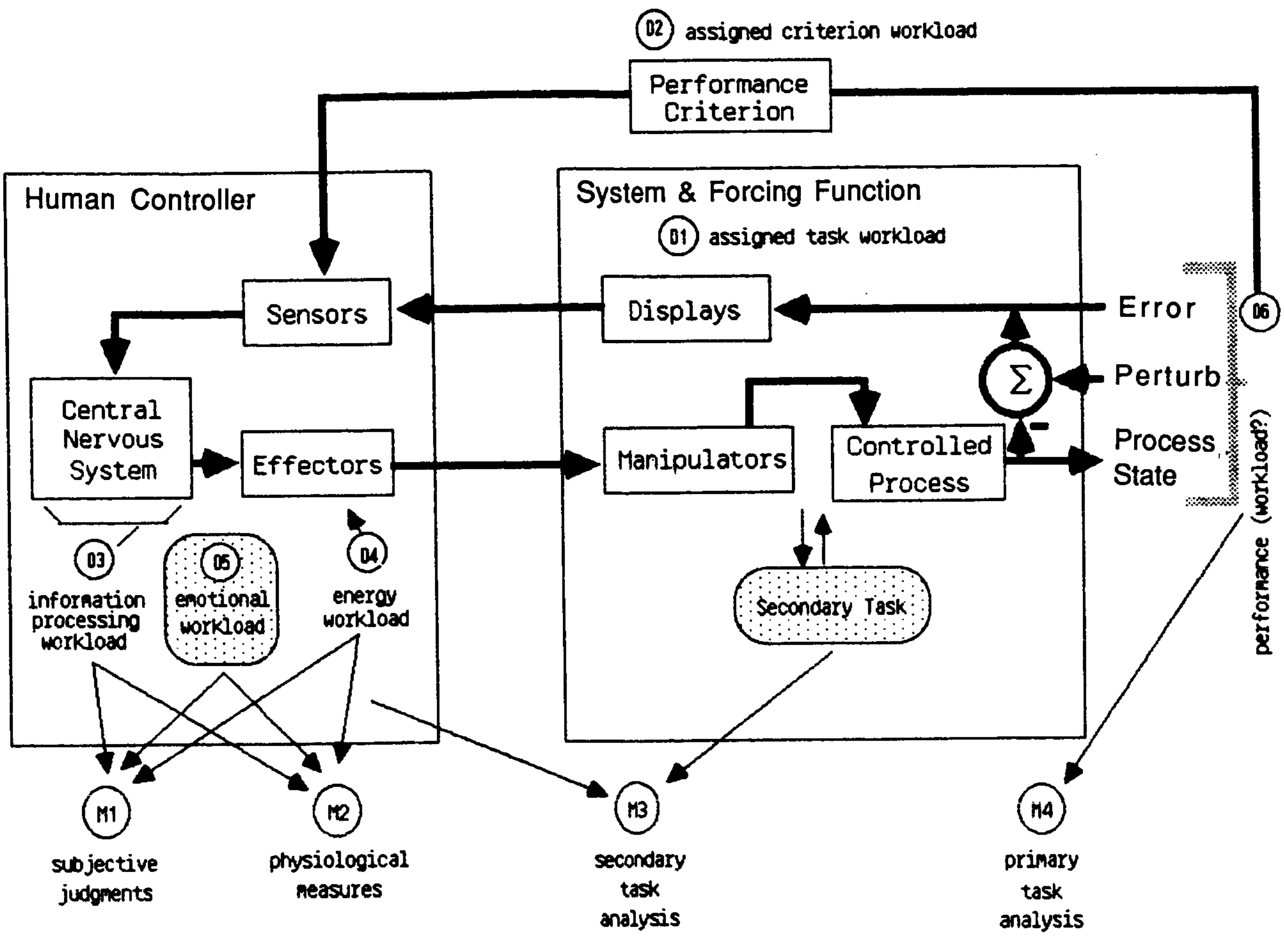
[17] **operator control of stimulus**

[18] **operator control of response**

[19] **reaction-time/feedback-lag relationship** - the ratio of the intervals defined by the (reaction) time from stimulus initiation to response initiation and the (feedback-lag) time from response initiation to feedback initiation

[20] **feedback** - how quickly feedback occurs once the response is made and is, thus, defined as a function of [19]

[21] **decision-making** - the multiplicity of choice-nodes, where the operator must decide which of several potential steps should be done next



(adapted from Sheridan & Stassen, Ref. 1, pg 242)

Figure 1. Alternative Workload Definitions.

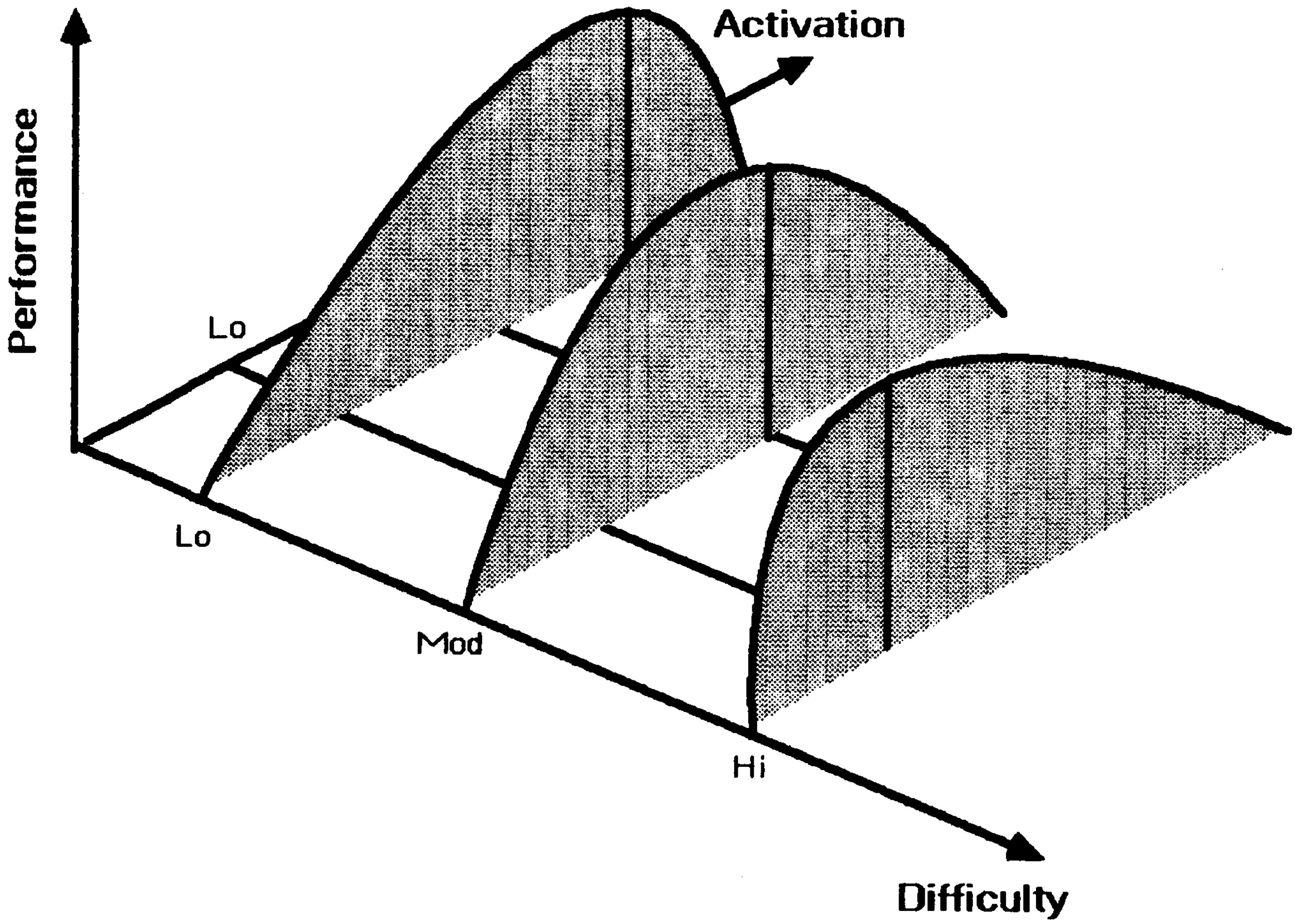


Figure 2. The Yerkes - Dodson Principle.

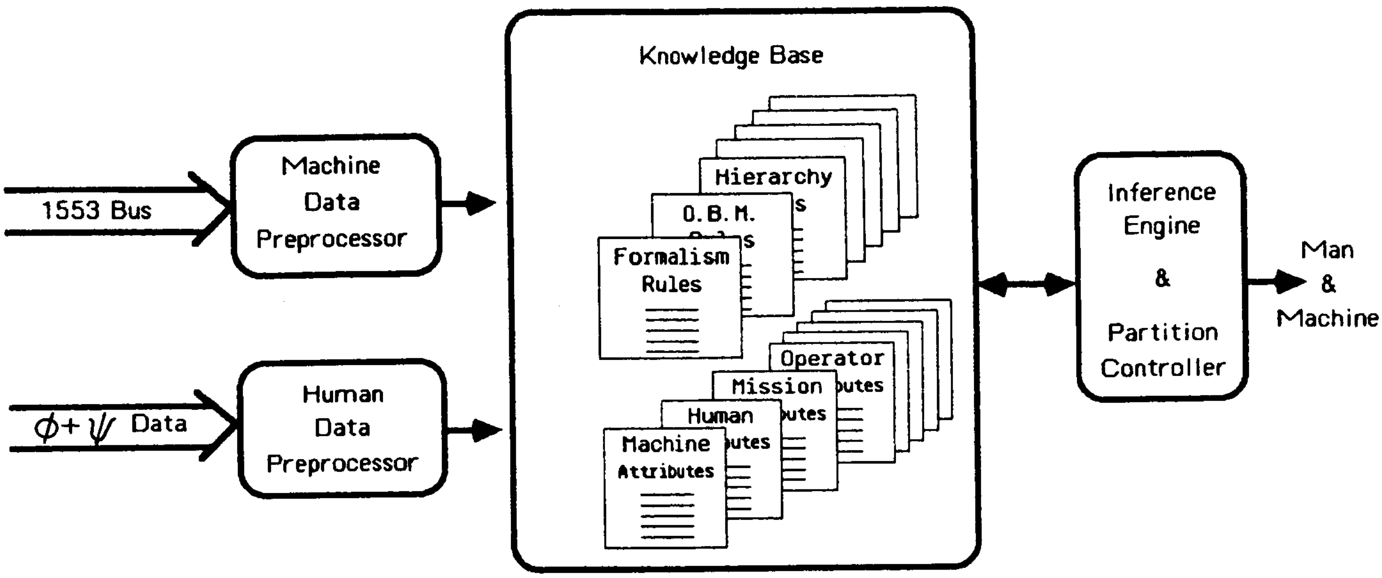


Figure 3. Metacontrol Unit Diagram.

VAGAL TONE AS AN INDEX OF MENTAL STATE

Stephen W. Porges
University of Maryland
College Park, Maryland

Overview.

Since autonomic processes such as heart rate are neurally mediated, it has been proposed that monitoring these variables will provide sensitive indicators of central nervous system status. Thus, many researchers have proposed that the neurally mediated oscillations in the heart rate pattern reflect a variety of mental states, including stress, emotion, consciousness or alertness, and attention. This paper will focus on the utility of monitoring oscillations in the heart rate pattern as a "window to the brain" and an index of general central nervous system status.

Heart rate in a healthy alert adult is not steady. The pattern of heart rate reflects the continuous feedback between the central nervous system and the peripheral autonomic receptors. The feedback produces phasic increases and decreases in neural efferent output via the vagus to the heart (ref. 1). In most situations like other measures of homeostatic function, the greater the range of the phasic increases and decreases, the "healthier" the individual. For example, with the aging process or with severe stress, there is an attenuation of the range of homeostatic function. Paralleling this process is a reduction in heart rate variability (ref. 2).

Thus, the efficiency of neural control may be manifested in rhythmic physiological variability and may portray the status of the individual and the individual's capacity and range to behave. In other terms, the greater the "organized" rhythmic physiological variability, the greater the range of behavior. Individuals with attenuated physiological variability, would then exhibit a lack of behavioral flexibility in response to environmental demands.

Although average heart rate seems to be a relative accurate index of metabolic activity, the topography of the heart rate pattern provides additional information regarding the continuous neural feedback between the cardiovascular system and the higher central nervous system structures. The spectral decomposition of the heart rate pattern identifies reliable oscillations at the respiratory frequency, at approximately .1 Hz hypothesized to reflect blood pressure feedback (e.g., Traube-Hering-Mayer wave,ref. 3), and at slower frequencies presumed to reflect thermoregulatory processes.

Heart rate variability is a complex and often ambiguous construct. It has numerous mediators. The same level of heart rate variability can be mediated by a variety of combinations of neural and extra-neural influences. Therefore, our research has focused on respiratory sinus arrhythmia, the one oscillation in the heart rate pattern for which the physiological mechanisms are known.

It is possible to provide empirical evidence that the amplitude of respiratory sinus arrhythmia accurately maps into the efferent influence of the vagus nerve on the heart. It has been proposed that respiration, either by a central mechanism or via a peripheral feedback loop to medullary areas, phasically inhibits, or "gates" the source nuclei of the vagal cardio-inhibitory fibers (ref. 3). Maximal inhibition of vagal efferent output occurs during the mid to late inspiratory phase and maximal vagal efferent output occurs during the expiratory phase.

Recent research on neural pathways of vagal cardio-inhibitory neurons has demonstrated that the vagal cardio-inhibitory neurons show a respiratory-related pattern of discharge with the primary efferent action on the heart occurring during expiration (ref. 4). Data from electrophysiological studies have been so consistent that functional properties including bradycardia to neural stimulation, pulse rhythm, and firing primarily during expiration have been used to determine when a neuron is a vagal cardio-inhibitory neuron (ref. 5).

Given the above characteristics of vagal cardio-inhibitory neurons, a strong argument may be made that quantification of the amplitude of respiratory sinus arrhythmia provides an accurate index of cardiac vagal tone. Since the vagal cardio-inhibitory neurons, by definition, slow the heart rate and exhibit a respiratory frequency, the impact on heart rate should be slowing during the expiratory phase of respiration. The greater the vagal efferent output to the heart, the greater the slowing of heart rate during expiration. Thus, respiratory sinus arrhythmia is a peripheral manifestation of the influence of the vagal cardio-inhibitory neurons on the heart (i.e., cardiac vagal tone).

Physiological model.

Vagal tone is quantified by measuring the spontaneous rhythmic heart rate changes associated with respiratory activity. Functionally, the sensory information is transmitted to the respiratory control area of the medulla from the stretch receptors in the lungs - monitoring inhalation and exhalation - as well as information from the chemoreceptors in the cardiovascular system reflecting blood gas composition levels of oxygen and carbon dioxide. This information "tunes" the medullary respiratory drive frequency.

The respiratory center influences the output of the vagus as it conveys neural information to the heart. The vagal efferents are modulated by the respiratory center, producing an attenuation of vagal efferent influences to the heart during inspiration, and a reinstatement of vagal efferent influences to the heart during expiration. Thus, the

phenomenon is known as respiratory sinus arrhythmia. The amplitude of respiratory sinus arrhythmia is not constant, but reflects higher brain influences which directly inhibit or stimulate the cells of origin of the vagus. Changes in the amplitude of respiratory sinus arrhythmia can be observed in studies of sustained attention, stress, anesthesia, sleep state, and in response to pharmacological treatments which depress the central nervous system. During many of these conditions, the respiratory parameters remain relatively constant.

The Vagal Tone Measure.

Assessment of vagal tone necessitates accurate quantification of the amplitude of respiratory sinus arrhythmia. Only the component of heart rate variance associated with respiratory sinus arrhythmia can be both "physiologically" and "empirically" related to vagal influences to the heart. The most sensitive measures of vagal tone must be based upon these constraints. We have developed a time series approach which accurately extracts from the complex heart rate pattern the amplitude of respiratory sinus arrhythmia. This measure has been labeled V to emphasize it is a measure of vagal tone.

This procedure solves many of the problems associated with employing time series statistics to study physiological processes. These problems include non-stationarity, aperiodic influences, and the fact that even when physiological processes are periodic, such as breathing and respiratory sinus arrhythmia, they are not perfect sine waves. The method includes a series of mathematically derived steps designed to enhance the study of periodic processes. Information associated with sampling rate, heart rate, and breathing rate need to be known and are incorporated in the algorithms (ref. 6). The methods are based upon knowledge of physiology and statistics. Misunderstanding of the method, either from a statistical or physiological dimension, may result in an inappropriate application and uninterpretable data.

Other estimates of vagal influence, such as measures of total heart rate variability or mean successive differences, often reflect interesting relationships with health status and behavior. However, these measures are less sensitive to manipulations of vagal control and are less consistent in demonstrating relationships with situational and physiological variables. Moreover, these measures are confounded by both physiological constraints (e.g., non-vagal influences) and statistical abberations (e.g., the sampling rate and the average heart rate influence the components of heart rate variability assessed with measures of heart rate variability that incorporate a successive difference approach). In many situations all measures of heart rate variability may be highly correlated, however, it can be demonstrated that the vagal tone measure is more sensitive to processes that can be physiologically linked to changes in parasympathetic tone.

These findings do not negate the importance of observations that global measures of heart rate variability are frequently related to mental states and clinical status. Rather, these points argue that global measures of heart rate variability are "composite" measures which can be

obtained through a variety of combinations of component influences on heart rate variability (such as movement, blood pressure feedback, respiratory sinus arrhythmia, and thermoregulatory influences). Therefore, it is impossible to make a strong statement regarding the specific physiological mechanisms mediating these relationships.

Validation studies.

To validate the vagal tone measure, a number of studies have demonstrated its sensitivity to manipulations of cardiac vagal tone (ref. 1). Our research has demonstrated that stimulation of the aortic depressor nerve in the rabbit increased the amplitude of respiratory sinus arrhythmia (ref. 7). Stimulation of the aortic depressor nerve produces a baroreceptor reflex characterized by increased vagal inhibitory action on the heart. Vagal blockade with atropine removed the effect. Propranolol, a beta-adrenergic blocker, did not alter the magnitude of the evoked increase in the amplitude of respiratory sinus arrhythmia. The amplitude of respiratory sinus arrhythmia was evaluated during manipulations of the baroreceptor reflex in anesthetized cats (ref. 8). Hypertension, induced by infusion of nitroprusside, was used to inhibit cardiac vagal tone. The manipulations effectively produced state changes in blood pressure and reflexively influenced the cardio-inhibitory influence on the heart (i.e., vagal tone). Hypertension produced an increase in the amplitude of respiratory sinus arrhythmia. Hypotension produced a decrease in the amplitude of respiratory sinus arrhythmia. Specific autonomic contributions were assessed with administration of practolol (a beta-adrenergic blocker) and atropine.

Although the above studies were conducted in anesthetized preparations, we also have conducted research with alert and moving preparations. In a study with rats, phenylephrine increased, atropine abolished, and saline had no effect on the amplitude of respiratory sinus arrhythmia (ref. 9). In a study with alert adults, four treatment levels of atropine and a placebo control were administered (ref. 10, ref. 11). The data demonstrated that the vagal blockade was monotonically related to the amplitude of respiratory sinus arrhythmia. Moreover, respiratory sinus arrhythmia was more sensitive to vagal blockade than heart rate (the change in heart rate in response to atropine is often used a criterion measure of vagal tone).

Sustained attention.

In a number of studies (ref. 12, ref. 13, ref. 14), heart rate variability was evaluated during a variety of attention demanding tasks. These studies demonstrated that independent of the direction of the heart rate change during the tasks, heart rate variability was consistently suppressed during sustained attention. Moreover, individuals with higher baselevel heart rate variability exhibited greater suppression of heart rate variability and performed better on reaction time tasks. These studies used a measure of overall heart rate variability and were conducted before the statistical procedures were developed to extract the amplitude of respiratory sinus arrhythmia.

Recently we have conducted research on the vagal tone measure during sustained attention. In this study physiological and performance measures were evaluated on 30 male and female students with a mean age of 20.2 years. The tasks were mental arithmetic and a tracking presented via an Atari Videoarcade system. Both tasks contained timers that were visibly displayed and that counted down while the subject tried to accumulate as many laps (in the tracking task) or points (in the arithmetic task).

For the tracking task, the subject was asked to race a video representation of a car around a track to make as many laps as possible in 60 sec. The task contained an element of uncontrollability. The task was designed so that there was a tradeoff between speed and accuracy. If the car went too fast for a certain period of time, the car would veer off course into the progress-impeding borders. The task required being able to control the speed while skillfully guiding the car to avoid the time-consuming border areas. Subjects were told that psychomotor skill was being assessed and that they would receive five cents for every completed lap. Subjects were given a practice session. The 60-second task was followed by a 60-second rest period referred in the figures as the "off-task" period. Each subject received three trials of an on-task/off-task sequence.

In the arithmetic task, five numbers between 1 and 9 were presented on the video screen for five seconds. A timer, displayed in the center of the screen, counted down while the subject tried to add the numbers together. Subjects responded by pressing a button. Subjects were rewarded for performance. Similar to the "race," the "sum" task was presented in three one-minute trials of an on-task/off-task sequence.

Collapsed across tasks heart period was shorter (faster heart rate) on task (748 msec) than off-task (850 msec); heart period variability was lower on-task (7.8) than off-task (8.4); respiratory frequency was faster on task (.33 hz) than off-task (.25 Hz); Vagal tone was lower on-task (7.4) than off-task (8.9); and the .1 Hz wave had lower amplitude on-task (7.3) than off-task (7.7).

A quantitative method of assessing the relative sensitivity of the above dependent variables to attention demands is to calculate eta or omega squared for the on-task/off-task effect. This procedure assesses the percentage of variance of the dependent variable mediated by the tasks. If the physiological variable is sensitive to the attention demands, it will be reflected in a greater percentage of the sums of squares associated with task relative to the total sums of squares in the analysis of variance table. The vagal tone index was the most sensitive of the physiological variables with an eta of .24 (i.e., 24% of the variance of vagal tone was mediated by the attention demanding tasks). The amplitude of the .1 Hz wave (i.e., Traube-Hering-Mayer wave), which has been reputed to be sensitive to sustained attention had the lowest eta and accounted for only 8% of the variance.

Flight performance and vagal tone.

The study by Dellinger, Taylor, and Porges (1987) (ref. 10) provides data on the relationship between changes in vagal tone and flight performance decrement. The injection of atropine resulted in significant performance decrements beginning at 1 hour post-injection and only minimal recovery by post-injection. In contrast, the decrement in vagal tone was almost instantaneous. The early physiological symptoms that occur prior to the performance decrements potentially could be used in bio-cybernetic system to allow the pilot to land safely.

Thus, although there is a parallel under the high doses of atropine (2.0 mg/75 kg, 4.0 mg/75 kg) between vagal tone and pilot performance, the time courses of the two classes of variables differ. The vagal tone measure reflects the immediate influence on the physiology although performance does not deteriorate for at least one hour, thus reflecting the pilot's ability to compensate.

Summary.

Other influences on central nervous system such as anesthesia, head trauma and sleep have been investigated. For example, inhalant anesthesia which blocks central nervous system monitoring of peripheral sensory information virtually eliminates vagal tone. The vagal tone monitored following head trauma in the intensive care unit predicts neurological outcome. Other studies have demonstrated that vagal tone shifts as a function of sleep state. In general the vagal tone index appears to monitor global states of the central nervous system and may be useful in screening the general state of pilots.

References.

1. Porges, S.W., McCabe, P., & Yongue, B.: Respiratory-Heart Rate Interactions: Psychophysiological Implications for Pathophysiology and Behavior. In Perspectives in Cardiovascular Psychophysiology, J. Cacioppo & R. Petty (eds.), 1982, pp. 223-264.

2. Porges, S.W.: Spontaneous Oscillations in Heart Rate: Potential Index of Stress. In Animal Stress, P.G. Moberg (ed.), 1985, pp. 97-111.

3. Hatch, J. P., Klatt, K., Porges, S. W., Schroeder-Sasheway, L., and Supik, J. D.: The Relation Between Rhythmic Cardiovascular Variability and Relativity to Orthostatic, Cognitive, and Cold Pressor Stress. Psychophysiology, vol. 23 (1), 1986, pp. 48-56.

4. Lopes, O.V., & Palmer, J.F.: Proposed Respiratory Gating Mechanism for Cardiac Slowing. Nature, vol. 264, 1976, pp. 454-456.

5. Gilbey, M.P., Jordan, D., Richter, D. W., & Spyder, K.M.: The Inspiratory Control of Vagal Cardio-inhibitory Neurons in the Cat. American Journal Physiology, 1983, vol. 343, pp. 57-58.

6. Jordon, D., Khalid, M.E.M., Schneiderman, N., & Spyer, K.M.: The Location and Properties of Preganglionic Vagal Cardiomotor Neurones in the Rabbit. Pfugers Archiv-European Journal of Phyaiology, vol. 395, 1982, pp. 244-250.

7. Porges, S. W.: Method and Apparatus for Evaluating Rhythmic Oscillations in Aperiodic Physiological Response Systems. Patent Number: 4,510,944. April 16, 1985.

8. McCabe, P.M., Yongue, B.G., Porges, S.W., & Ackles, P.K.: Changes in Heart Period, Heart Period Variability and a Spectral Analysis Estimate of Respiratory Sinus Arrhythmia During Aortic Nerve Stimulation in Rabbits. Psychophysiology, vol. 22, 1984, pp. 195-203.

9. McCabe, P.M., Yongue, B.G., Porges, S.W., & Ackles, P.K.: Changes in Heart Period, Heart Period Variability and a Spectral Analysis Estimate of Respiratory Sinus Arrhythmia in Response to Pharmacological Manipulations of the Baroreceptor Reflex in Cats. Psychophysiology, vol. 21, 1985, pp. 149-148.

10. Yongue, B.G., McCabe, P.M., Porges, S.W., Riviera, M., Kelley, S.L., & Ackles, P.K.: The Effects of Pharmacological Manipulations that Influence Vagal Control of the Heart on Heart Period, Heart-Period Variability, and Respiration in Rats. Psychophysiology, vol. 19, 1982, pp. 426-432.

11. Dellinger, J.A., Taylor, H.L., & Porges, S.W.: Atropine Sulfate Effects on Aviator Performance and on Respiratory-Heart Period Interactions. Aviation and Space Environment Medicine, vol. 54, 1987, pp. 1242-1253.

12. Porges, S.W.: Respiratory Sinus Arrhythmia: Physiological Basis, Quantitative Method, and Clinical Implications. In Cardiorespiratory and Cardiosomatic Psychophysiology. P. Grossman, K. Janssen, & D. Vaitl, (eds.), 1986, pp. 101-115.

13. Porges, S.W., & Raskin, D.C.: Respiratory and Heart Rate Components of Attention. Journal of Experimental Psychology, vol. 81, 1969, pp. 487-503.

14. Porges, S.W.: Heart Rate Variability and Deceleration as Indexes of Reaction Time. Journal of Experimental Psychology, vol. 92, 1972, pp. 103-110.

15. Walter, G.F., & Porges, S.W.: Heart Rate and Respiratory Responses as a Function of Task Difficulty: The Use of Discriminant Analysis in the Selection of Psychologically Sensitive Physiology Responses. Psychophysiology, vol. 13, 1976, pp. 563-571.

N88-23374 S4-53 140792 15P. NQ791920

CHALLENGES OF PHYSIOLOGICAL MONITORING IN A NAVY OPERATIONAL SETTING

Guy R. Banta, MSC,
USN
Naval Medical Research and Development Command
Bethesda, MD

The requirements to evaluate naval personnel performance and the challenges of quantifying the physiological requirements of varied work tasks are inherent in all naval communities (diving, marine amphibious assault, special warfare, surface, sub-surface, and flight). The objectives of such requirements include not only safety but the development of selection and performance standards to better conserve personnel and maintain combat readiness in the Navy.

The effect of naval operations with a mix of multiple interactive stressors, e.g., extended operations, G-loading, thermal loads, motion/disorientation, multi-cognitive tasks, hypo/hyperbaric exposure, etc., as well as disease and injury potential, requires the development of critical field usable assessment techniques. Data that can be obtained from such techniques in the varied operational settings form the basis for simulation of adverse field environments in the laboratory where medical capability (selection/retention) criteria, mission modeling, man-machine interface design, and performance enhancement techniques can be developed, studied, and eventually fleet implemented. Examples of current Navy R&D thrusts in field physiological monitoring include: (a) performance assessment of Navy divers and combat swimmers during extended water exposure, (b) event related potential (ERP) monitoring of surface ship and submarine sonar operators, (c) in-flight assessment of fatigue indices in the anti-submarine warfare community and (d) quantification of cardiac stress in fighter pilots during air-to-air combat maneuvers.

DIVING

The U. S. Navy diving community is principally composed of three types of divers: (1) the traditional "Hard Hat Diver" who is responsible for diving to great depth for such missions as ship slavage, (2) the Explosive Ordnance Diver/Swimmer who is principally a shallow water diver responsible for placement/removing/disposing of ordnance in areas such as harbors, beach assault fronts, etc. and (3) the Naval Special Warfare operator (better known as a Sea/Air/Land (SEAL) Special Commando or Frogman) who in addition to many land base commando responsibilities is responsible for shallow water diving/swimming that may require greater than 6 hours of continuous immersion in varied temperature waters.

In order for mission accomplishment (peace time as well as combat) varied physiological/psychological concerns are presented for the Navy diver: decompression sickness, fatigue, cardiac stress, equipment malfunction, and the most prevalent, hypothermia, each of which leads to performance decrement and a threat to life. Reasoning for physiological monitoring for the sake of identifying such decrement/occurrences are twofold, (1) as a means for team/individual warning signals for completion/termination of mission or for implementation of protection techniques and (2) as an R&D effort to better understand the physiological requirements and responses of diving operations so that improved enhancement techniques can be developed.

A recent diving R&D effort was conducted to evaluate diver thermal status and performance during cold water immersion wearing dry or wet suits. Activity level, diet, time of day for the dive, and water temperatures were varied. Core and skin temperature responses and heat flux-data were collected (Fig 1). The data acquisition system as described by Weinberg (Ref 1) included skin and rectal temperature transducers with a constant current Wheatsone bridge circuit located near the computer to convert the changes in thermistor resistance into voltage levels. This provided a high S/N* ratio in the electrically noisy hyperbaric chamber environment and allowed simple computer sampling. Heat flux was converted to a multivolt level signal by a sensor disc that contained integrally the thermistor. Thermal data obtained are being used to validate several models of diver thermoregulation to produce safe exposure, guidelines, and selection charts for thermal protection garments.

In 1976 a National Plan (Ref (2)) was published which addressed concern for diver safety and performance decrements and the importance of varied monitoring techniques in the operational setting. These techniques included voice communications, heart rate, and respiratory and thermal parameters. A 1978 workshop (Ref (3)) supported these issues and developed the following conclusions regarding physiological monitoring of the diver.

a. The importance of visual monitoring
b. That monitoring suffers from lack of adequate sensors
c. A majority of the difficulty is in interpreting what is monitored
d. There is a need to assess physiological response against time
e. There is a need to reference physiological variables against individual physiological profiles
f. There is a need for real time display of physiological variables for supervisor use.

---

*S/N (signal-to-noise ratio).

These conclusions have not changed dramatically since 1978.

Outside of a controlled laboratory setting, operational feasibility is the major limiting factor for good physiological monitoring in a diving setting. Whether data transfer is by cable or telemetry or pier/ship side monitoring equipment, diver "Jury/Rigging" is the norm (Figs (2) & (3). Use of physiological sensors are continually difficult due to environmental exposure, time required for attachment, acceptance for use (e.g., rectal thermometer) data transfer interference, and quick data interpretation (especially for the individual diver). As addressed in (Reference (3)), the "ideal" diver operational monitoring system is a system that will monitor physiological and environmental parameters by sensors built into the diving suit and equipment, with real time digital readout, error-free data link to a monitoring point capable of immediate analysis, comparison with the individual diving profile, and prediction of outcome.

ELECTROPHYSIOLOGICAL MONITORING

Navy sonar operators are common to both surface and subsurface ships. Their tasks mandate a continued vigilance on sonar screens for any indication of nearby vessels. The nature of this job includes extensive information processing, frequent dual tasks, and a requirement for unobstructed attention and performance for sustained periods of time. We know that the human information processing system is limited in its capacity to handle multiple inputs and is subject to diverted attention and fatigue even in the best of trained operators. In a naval combat environment such deviation could prove disasterous. A current hypothesis is that decrements in neuroelectrophysiological components, which have been found to be highly correlated to attentional mechanisms (Ref (4)), can be detected prior to the onset of actual performance decrement. This theory provides a potential for R&D laboratory assessment and development of a performance monitoring tool for shipboard use. In a laboratory setting collection of neuroelectric responses during varied simulated tasks can be utilized to determine several possible performance counter-degradation techniques e.g., improved sleep management doctrine and crew rest/work cyles, provision of pharmacological aids, or as a selection tool for identifying the best performers. A recent study (Ref (5)) directed at exploring the feasibility of neuroelectric monitoring of sonar operators used highly trained sonar operators and investigated signal detection and signal recognition in a simulated sonar task. During presentation of sonar targets event-related potentials (ERPs) were recorded from a number of electrode sites over a 1750 msec recording period (Fig 4). Results revealed that several ERP components were significally related to some aspect of detection and/or recognition. Figure (5) demonstrates results for targets correctly recognized. A positive response is demonstrated by a

downward deflection of the large P300 wave forms. These data identified that ERP components may be useful in evaluating detection and recognition performance. Utilization of a neuroelectric monitoring system aboard ship as a means of identifying deviation of attention/recognition (that is decrement of performance) in sonar operators has been contemplated. To date, no such system has been implemented. The following issues, similar to those expressed about physiological monitoring in the diving community, would have to be addressed before implementing such a monitoring system: (1) compliance, (2) means of instant data analysis and display of results, (3) durability and upkeep of equipment in an at-sea environment and (4) supervisory monitoring.

IN-FLIGHT MONITORING

Attempts at in-flight monitoring of physiological responses are not known. In the 1960s NASA supported an effort in which aircrew physiological response was recorded during flight operations over Vietnam. Although the number of subjects and physiological variables was limited, it was surprising to discover that carrier launch and recovery operations were more stressful (physiologically) than combat (Refs 6, 7). With aircraft design progression that has resulted in development of higher thrust-to-weight rates, reduced wing loads, and maneuverability that expose aircrew to greater than 10 Gs for sustained periods of time, an extensive need for better understanding of physiological response has been created. Proper assessment of the magnitude of response (physiologically) to aviation task loading is by in-flight monitoring of selected physiological responses.

AIRCREW FATIGUE

Based on a Chief of Naval Operation's guidance, an effort to investigate fatigue in the Navy's Anti-Submarine (ASW) patrol community was initiated. The directions were to assess whether fatigue exists, and if found, determine the effect on flight and mission performance.

Of the many definitions of fatigue, "task-induced" fatigue appears to best fit the Navy's operational environment. It is a fatigue produced by long hours of work in a taxing environment where loss of efficiency is attributed to both physiological and psychological factors. The concern about fatigue during sustained flight operations is one of both safety of flight and mission completeness.

The literature reveals an extensive amount of work in the area of fatigue and the varied methods of monitoring fatigue in aviation environments, dating back as early as the 1930s. Emphasis has most often been directed at cargo and transport aircraft conducting trans-oceanic, multihour, and multicrew

flights (Refs 8, 9, 10) with a few studies addressing tactical jet flights during carrier operations (Refs 11, 12) and even fewer in the ASW community (Refs 13, 14). It is quite clear that the level of fatigue can vary by job task. In fact, with dependency on operational tempo and lack of in-flight relief, fatigue may at times be more prevalent in the non-flight deck crew.

Physiological parameters frequently identified as indices of fatigue are: (a) varied endocrine responses such as blood catecholamines and urine levels of cortisol, 17-hydroxycorticosteroids, sodium/potassium concentration, urea, and creatine levels, (b) heart rate and electrocardiographic (ECG) response, (c) electromyogram (EMG) for assessment of muscle tension, (d) body temperature, and (e) muscular strength.

Routine collection of blood during in-flight periods (even in a multicrew sized aircraft) for blood levels of fatigue indices is (1) not easily accepted by the crew and (2) is not feasible because of a need to centrifuge blood samples and freeze them immediately. Therefore, urine samples collected during or at pre-post flight times and stored for later analysis have been a reasonable approach. Monitoring of variables such as heart rate, EKG, and EMG have been assessed in previous studies using varied gear driven tape recorders (Ref 15), telemetry systems (Ref 16), and limited solid state recording devices (Ref 17). In large multicrew aircraft such monitoring by medical personnel can be accomplished using more laboratory type monitors such as magnetic tape recorders, strip charts, etc.

A recent series of studies addressing in-flight fatigue in Navy ASW P-3 aircraft during overseas deployment have been conducted. Each aircrew member of a selected P-3 crew (usually 9 crewmen) was assessed physiologically (blood chemistries, muscular strength, aerobic fitness, etc.) prior to a 6 month deployment. While on deployment selected physiological parameters, as described above, were monitored/collected during actual multiple ASW flights (Fig 6). Following return from deployment, performance assessment data as conducted during pre-deployment were again collected. As expected, analysis of the data revealed significant changes in fatigue indices during flight, as well as, over a 6 month deployment period (Ref 14).

The process of physiologically monitoring ASW aircrewmen during operational flights proved to be very difficult with the loss or non-acceptance of much of the data. Timing of data collection could not be controlled (mission operation interference), equipment idiosyncratic responses could not be easily corrected during flight, and efficient processing of blood and urine samples in field conditions versus sophisticated laboratory environments stimulated considerable concern about the size of the "standard deviation" in the data.

IN-FLIGHT CARDIAC STRESS

Naval/Marine Corps aviators flying high performance aircraft are exposed to frequent and repeated environmental and operational tasks, e.g., excessive $\pm$ G loading, high oxygen demands, high temperatures, barometric pressure changes, disorientation, extreme visual tracking requirements, etc. These tasks pose physiological stresses/demands that can degrade performance. Studies are beginning to show that subjects whose energy requirements, metabolic activity, thermal, and cardiopulmonary states are least disturbed by the stress of high performance flight will perform best and become least fatigued during repeated aerial task loading (Refs 18, 19). During aerial combat, aviators use a spectrum of G levels varying in a continuous manner called the Aerial Combat Maneuver (ACM). The G envelope in which they fly may range from -3 to +10 G_z. Although ACM is a common flight environment, there is limited knowledge of the human tolerances to this environment.

Concerns about the effect of physical fitness on cardiac stress during high performance flight initiated a study in which G-load and heart rate response were collected during air combat maneuver (ACM) training flights. Several naval fighter pilots flying ACM training flights on a Tactical Air Combat Training System (TACTS) range were used as subjects. Heart rate response was collected every 2.5 seconds during flight by an eight channel sold state recording device (Fig 7). The monitor was attached by 3 ECG chest leads and was carried in the aviator's flight suit pocket. Aircraft flight responses were collected by a telemetry device attached to the aircraft wing which transmitted real-time data to a ground based computer system. Results demonstrated significant changes in heart rate during all phases of the flight profile (Fig 8) with inverse relationship between heart rate response and level of fitness.

The significant difficulty encountered with physiologic monitoring in this operational setting was the necessity to collect continous heart rate response due to an inability to start/stop the monitor during flight. This necessitated extensive postflight data analysis. Additionally, a difficulty of extreme interest was the lack of accurate time-phasing of heart rate response with specific aircraft maneuvers because of the noncommunication between the two monitoring systems. For the most accurate human response data collection in an operational setting, there must be provisions for sequencing of a given response to the operational task in order to best determine "cause and effect".

IN-FLIGHT MONITORING DEVICES

In the latter part of the 1960's, the U.S. Navy developed an in-flight monitor called the Bio-Pack (Ref 15). The Bio-Pack consisted of a small gear driven 8 channel reel to reel recorder with a recording time of 40 minutes. Physiological variables monitored were ECG, body temperature, voice, cockpit acceleration, and temperature. The Bio-Pack was carried by the aviator on his knee board or map case with minimal interference. In 1975 the U. S. Air Force expanded the Bio-Pack by increasing its recording time and making it a gear driven cassette tape recorder. By the late 70's the Navy again expanded the recorder by developing a data analysis program and expanded the original 8 channel recorder to a 32 channel recorder and changed the name to In-flight Physiological Data Acquisition System (IFPDAS). When these gear driven data recorders were used in tactical jet aircraft, the prevalent difficulty was periodic tape speed fluctuations due to periods of excessive acceleration force on the tape recorder. Other operational constraints have been intermittent signals from the transducers, fixed data sampling rates, excessive maintenance efforts, and incompatibility between the microprocessor and other data processing units. The goal of in-flight physiological monitoring has been to accurately monitor and collect as many physiological responses and environmental data points during flight as possible with minimal interference to the pilot. A monitoring device capable of doing this must be small in size and self contained with multiple channels and expanded memory capability. Using today's state of the art in microprocessing and memory technology, the Navy has recently designed a Solid State Physiological Inflight Data Recorder (SSPIDR) for use in aeromedical flight test operations (Fig 9). "The SSPIDR incorporates analog to digital physiological signal conversion, a Motorola 68000/6/32 bit microprocessor, 512K x 16 bit memory, and battery power supply in a 13 x 15 x 5 cm package which fits in an aircrewman's survival vest. Available data channels include three electrocardiogram, one electroencephalogram, one respiratory rate, two electrooculogram, three linear acceleration, eight temperature, one pressure, and a digital event marker. On-board software permits variable sampling rates and gain. The sampling rate and amplification may be changed according to predefined flight or physiological conditions" (Ref 20). The Navy has also recently initiated a contract with a bioengineering company to design and fabricate a miniature O_2/CO_2 transducer for use in the SSPIDR.

Outyear plans for the SSPIDR include monitoring of physiological responses during wearing of chemical defense ensemble, sustained operations, use of varied pharmacological agents, parachuting, impact/acceleration, and multiple cognitive task regimes. The device will also be made available for other Navy community uses as applicable. The SSPIDR, however, is not yet designed for water immersion.

CONCLUSIONS

Physiological monitoring is necessary to identify the physiological requirements of operational tasks, how well the individual performs within those tasks, and the success or failure of the man-machine interface. In the Navy, operational settings are numerous and unique. The challenges of physiological monitoring in the varied operational settings are also extremely numerous. The general difficulties of monitoring are, however, most likely commonplace among the settings and among the services. These challenges include, but are not limited to, environmental extremes, acceptance of use by test subjects, data transfer, data interpretation, and capability of relating collected data to valid operational relevant criterion measures.

REFERENCES

1. Weinberg, R. P.: A Data Acquisition System in Thermal Physiology. Proceedings IEEE COMPCON Fall Conference on Small Computer Revolution, September 1984.

2. Galerine, A., Ackles, K., Duff, C., Hall, D., Linderoth, L. S. and Spencer M.: Monitoring Techniques in National Plan for the Safety and Health of Divers in Their Quest for Subsea Energy. M. W. Beckett, ED., Undersea Medical Society, 1976.

3. Technical Report #WS:78-2. The Sixteenth Undersea Medical Society Workshop, Monitoring Vital Signs in the Diver, M. W. Beckett, ED., Undersea Medical Society, 1978.

4. Naatanen, R.: Processing Negativity: An Evoked-Potential Reflection of Selective Attention. Psychological Bulletin: 92(3), 605-640, 1982.

5. Kobus, D. A., Beeler, M. J., and Stashower, K.: Electro-physiologoical Effects of Experience During an Auditory Task. Naval Health Research Center Tech Report #87-2, 1987.

6. Lewis, C. E., Jones, W. L., Austin, F. and Roman, J. Flight Rsearch Program: IX Medical Monitoring of Carrier Pilots in Combat - II. Aerospace Med 38:581-592, 1967.

7. Roman, J., Older, H. and Jones, W. L.: Flight Research Program: VII. Medical Monitoring of Navy Carrier Pilots in Combat. Aerospace Med 38:133-139, 1967.

8. Hale, H. B., Anderson, C. A., Williams, E. W. and Tanne, E.: Endocrine-Metabolic Effects of Unusually Long or Frequent Flying Missions in C-130 E or C-135B Aircraft. Aerospace Med 39 (6): 561-570, 1968.

9. Lyman, E. G., and Orlady, H. W.: Fatigue and Associated Performance Decrements in Air Transport Operations. NASA-CR 166167, 1981.

10. Storm, W. F.: E-4B Crew Fatigue Associated with 30-Hour IOI and E Missions. USAFSAM San Antonio Tech Report 80-40, 1980.

11. Kramer, E. F., Hale, H. B. and Williams, E. W.: Physiological Effects of An 18 Hour Flight in F-4c Aircraft. Aerospace Med 37 (11): 1095-1098, 1966.

12. Marchbanks, V. H., Hale, H. B., and Ellis, J. P.: Stress Responses of Pilots Flying 6 Hours Over Water Missions in F-100 and F-104 Aircraft. Aerospace Med 34: 15-18, 1963.

13. Banta, G. R. and Bory, A.: Fatigue in Navy Patrol Community Sustained Flight Operations (Abstract). Aviation, Space and Environmental Medicine, 1985.

14. DeJohn, C. A., Murdock, D. M., Parker, J. B. and Banta, G. R.: Effects of Sustained Flight Operations in Naval Aircraft. Aviation, Space, and Environmental Medicine 57 (5): 487, 1986.

15. Horrigan, D. J. Jr., Martin, J. D. and Englorecht, W. J.: A Self-Contained System for In-Flight Acquisition of Physiological Data. Naval Missile Test Center TR-70-61 (U). 29 October 1970.

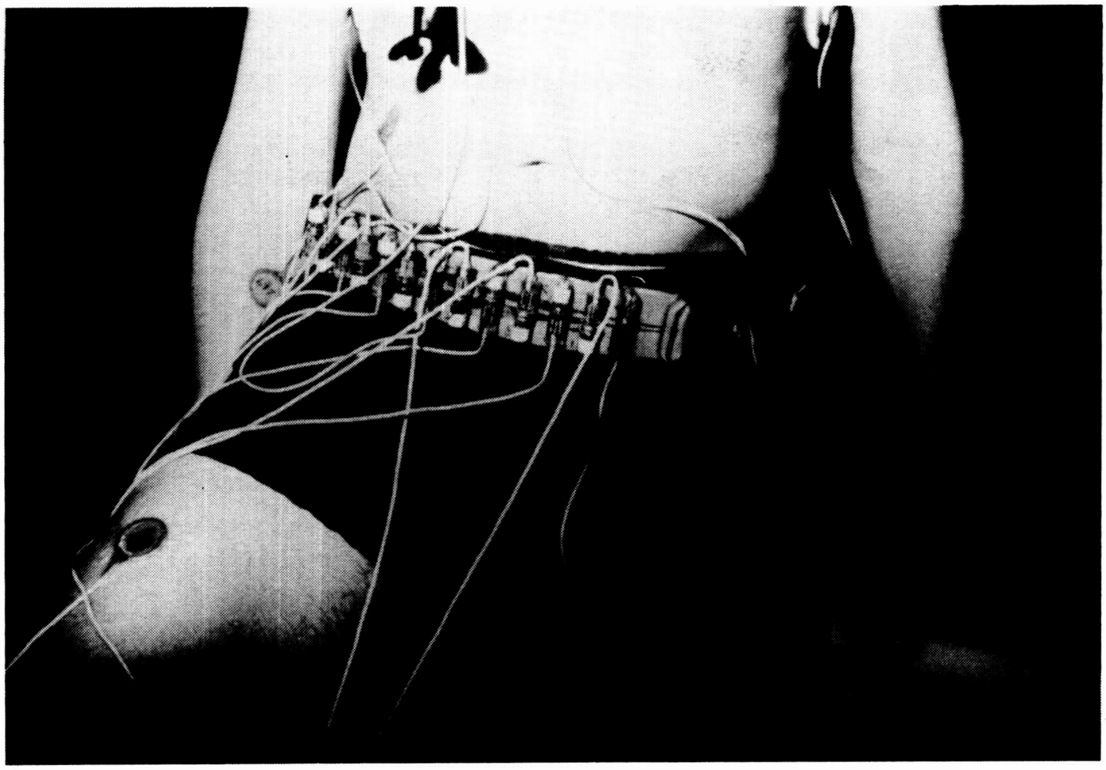
16. Nunnley, S. A. and James, G. R.: Cockpit Thermal Conditions and Crew Skin Temperatures Measured in Flight. Aviation, Space and Environmental Medicine 48: 44-47, 1976.

17. Banta, G. R. and Grissett, J. D.: Relationship of Cariopulmonary Fitness to Flight Performance in Tactical Aviation. Naval Aerospace Medical Research Laboratory Tech Report #1329, January 1987.

18. Burton, R. R.: Human Responses to Repeated High G Simulated Aerial Combat Maneuvers. Aviation Space Environmental Med 51(11): 1185-1192, 1980.

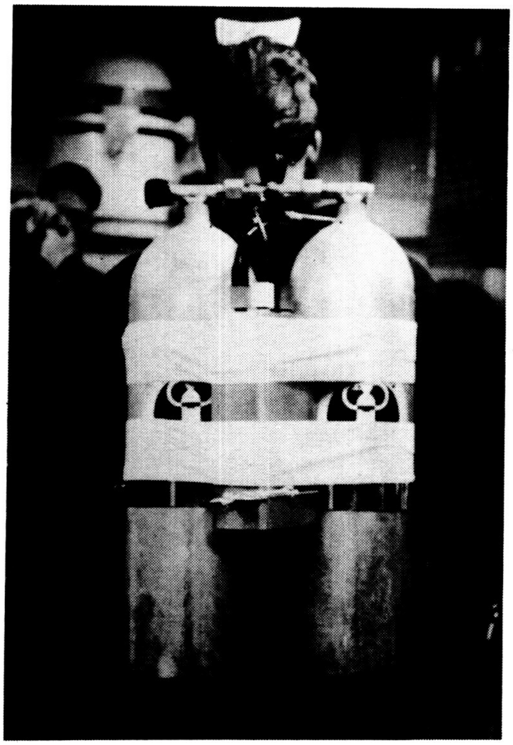
19. Roman, J. A.: Cardiorespiratory Functioning In-Flight. Aerospace Med Apr: 322-337, 1963.

20. Call, D. W. and Miller, W. F.: A Solid State Physiological Inflight Data Recorder Aviation Space and Environmental Med 58(5): 508, 1987.



1. Diver Heat Flux and Thermal Monitoring System

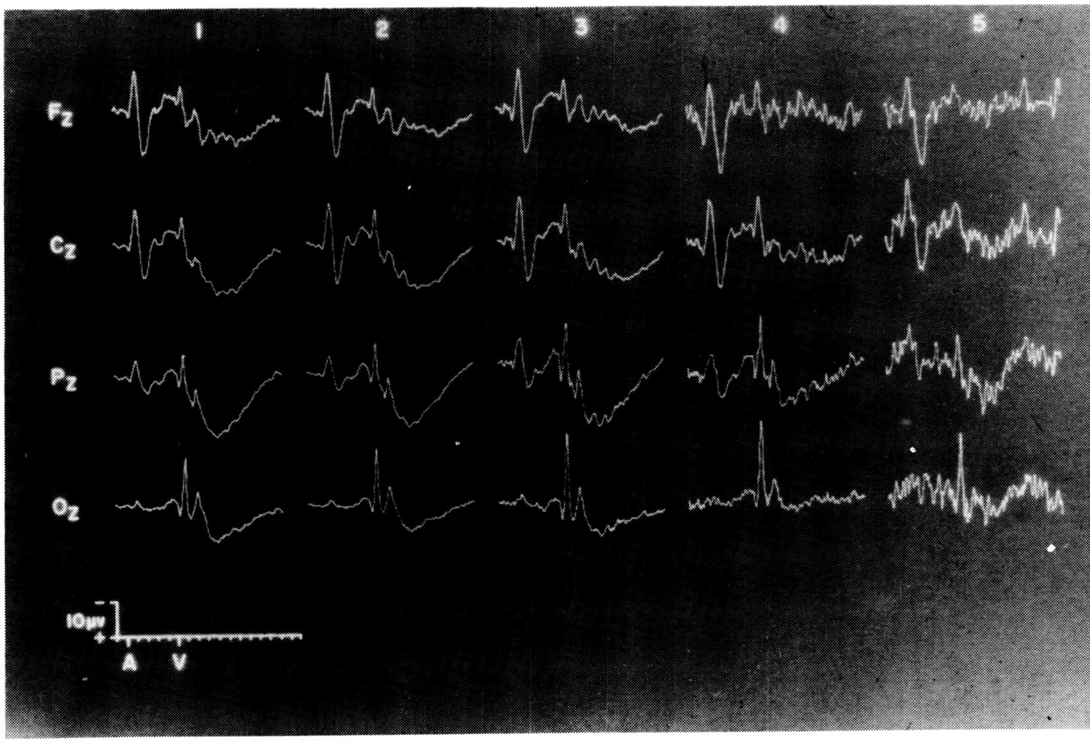
2. Attachment of underwater physiological equipment



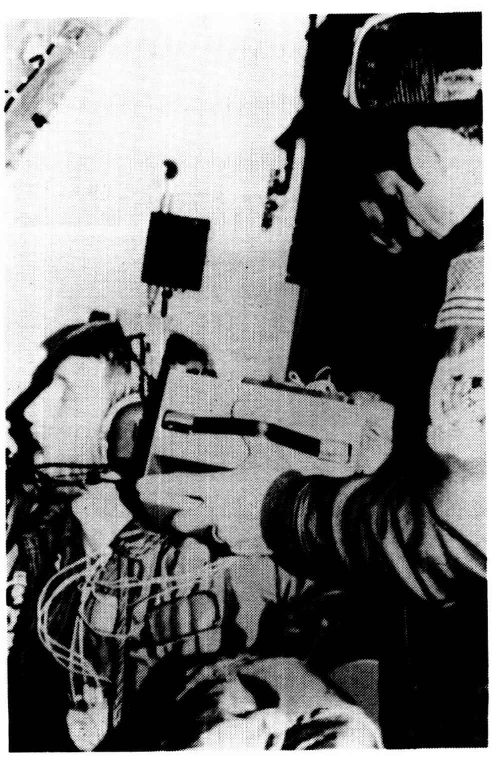
3. Navy Diver with a physiological data acquisition system taped to air tanks

4. Attachment of event related potential (ERP) recording electrodes on a U. S. Navy sonarman

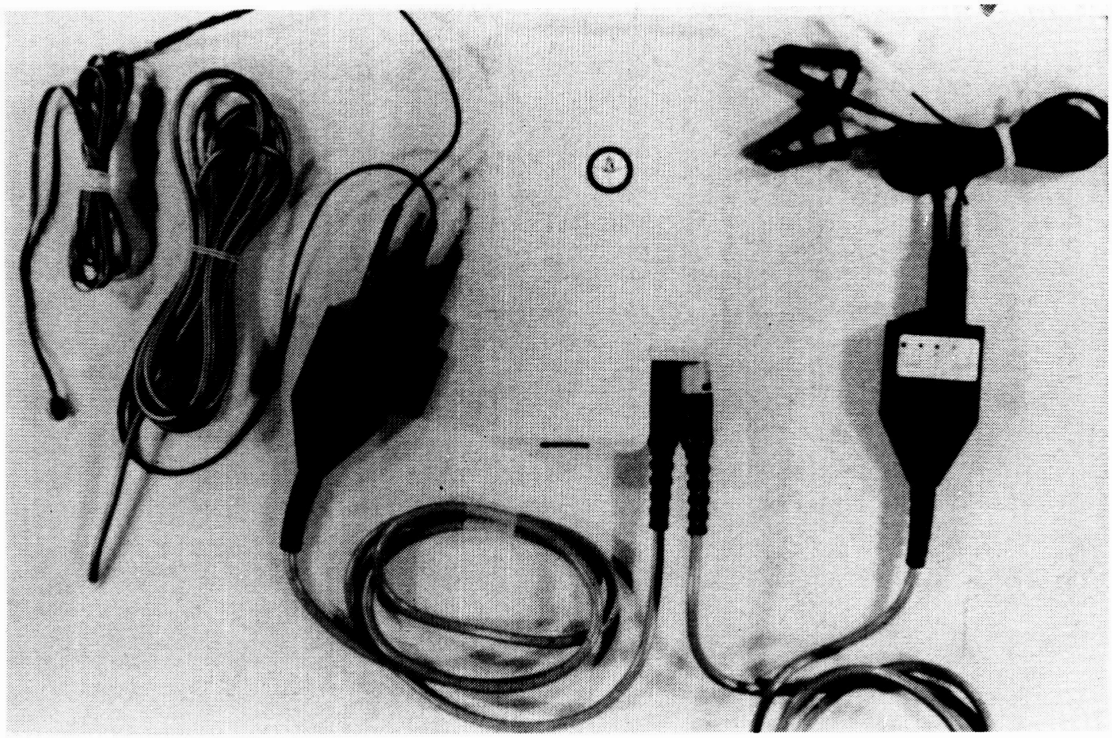
ORIGINAL PAGE IS
OF POOR QUALITY



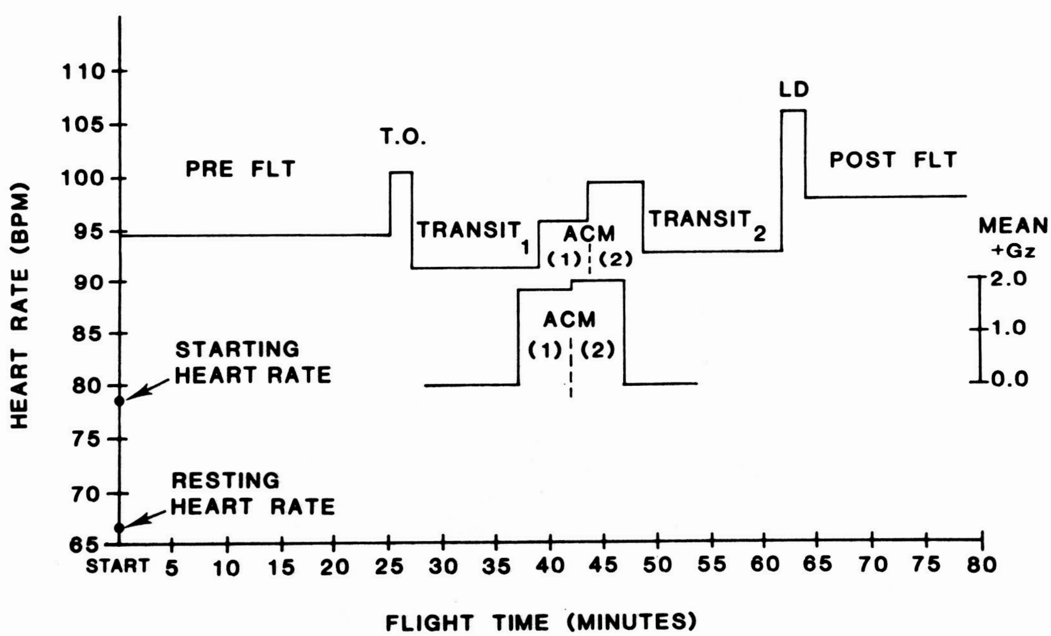
5. Event related potential (ERP) recordings from four skull sites: F_2, C_2, P_2, 0_2. Positive response to target recognition is demonstrated by downward deflection of the large P300 wave forms



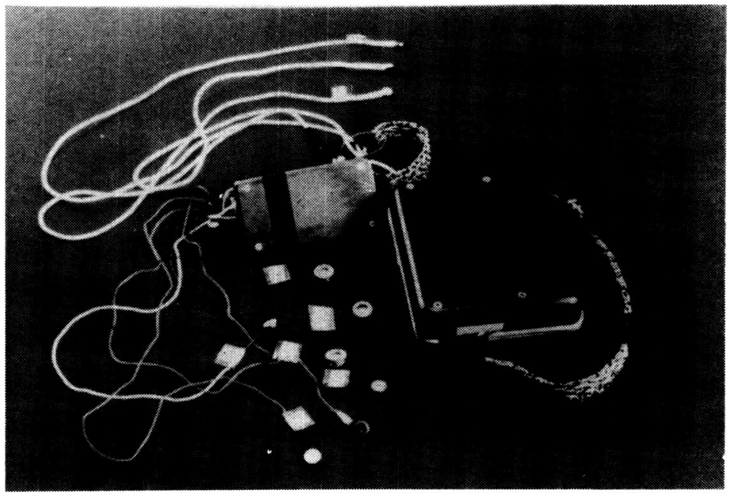
6. In-flight physiological monitoring of an antisubmarine warfare patrol aircraft aircrewman



7. Eight-channel solid-state recording device (Trade name: Vitalog (Vitalog Corp))



8. Mean heart rate and $+G_z$ response during Tactical Air Combat Training System (TACTS) Range ACM flights for 11 aviators during 23 flights. PRE flt = preflight; T.O. = takeoff; transit = flight to TACTS range. ACM 1 and 2 = individual ACM events (flights); transits = return flight to base; LD = landing, Post Flt = postflight (Banta et. al., Naval Aerospace Medical Research Labortory, TR # 1329, January 1987)



9. Solid-State Physiological Inflight Data Recorder (SSPIDR) used for aeromedical flight test operations

N88-23375 55-53 140793 16P

PREDICTING OPERATOR WORKLOAD DURING SYSTEM DESIGN

Theodore B. Aldrich and Sandra M. Szabo
Anacapa Sciences, Inc.
Fort Rucker, Alabama

AU 848381

INTRODUCTION

Current Army policy requires that human capabilities and limitations be addressed during the conceptual phase of new weapon systems development. In furtherance of this policy, Anacapa Sciences, Inc. researchers, under contract to the U.S. Army Research Institute Aviation Research and Development Activity (ARIARDA), developed a methodology to predict aviator workload in advance of aircraft system design. The methodology features models that predict workload under varying automation configurations for both single- and multi-crew system designs. This paper (a) describes the methodology for developing and exercising the workload prediction models and (b) presents flight simulator-based research plans for validating the workload predictions yielded by the models.

THE WORKLOAD PREDICTION METHODOLOGY

Background

The Army's Air/Land Battle 2000 scenario represents a high-threat environment that will place heavy workload demands on Army aviators. Accordingly, future aircraft systems are being developed with advanced technology designed to automate many of the functions traditionally performed by crew members. Examples of the advanced technology include:

- an increased number of sensors and target acquisition aids
- improved navigation and communication systems
- advanced crew station design features
- improved flight controls
- extraordinary avionics reliability
- subsystems that are automatically reconfigured if components fail

Although advanced technology is typically designed to reduce aviator workload, the tasks required to use the

technology may actually increase workload in some instances. For example, technology designed to reduce an aviator's need to maintain physical control of system functions often increases the aviator's role as a systems monitor or problem solver. Consequently, while psychomotor workload demands are decreased, sensory and cognitive attentional demands are increased.

The development of new and improved aircraft systems also presents problems in the prediction and assessment of operator workload. Metrics that are appropriate for analyzing physical workload are inadequate for assessing sensory and cognitive workload. Accordingly, workload research has shifted from a focus on physical effort required to perform a task to an emphasis on the attentional demand associated with the sensory, cognitive, and psychomotor workload components of the tasks. The workload prediction methodology developed by ARIARDA and Anacapa researchers operationally defines workload in terms of attentional demand. Consequently, the methodology is designed to measure "mental state" associated with task performance.

The workload prediction methodology was developed in response to a request for research support from the Army's Aviation Systems Command (AVSCOM) Program Office charged with the development of a new multipurpose, lightweight helicopter, designated the LHX. A detailed description of the manner in which the methodology was developed and applied to the LHX is presented in reference 1.

The original LHX workload prediction methodology currently is being refined during analyses of three additional Army helicopter systems and one advanced-technology crew station for an experimental research flight simulator. The four additional analyses are:

- a baseline analysis for the AH-64A, Apache, prior to predicting crew workload in a proposed AH-64B configuration (ref. 2)

- a baseline analysis for the UH-60A, Blackhawk, prior to predicting crew workload in a redesigned MH-60X configuration (ref. 3)

- a baseline analysis for the CH-47, Chinook, prior to predicting crew workload in a redesigned MH-47E configuration

- a baseline analysis for an advanced technology LHX-type crew station for the Crew Station Research and Development Office (CSRDO) at NASA Ames, prior to predicting crew workload in high-fidelity flight simulation experiments

In applying the methodology to the aircraft and flight simulator systems, three major phases of research must be performed:

- conduct mission/task analyses of critical mission segments and assign estimates of workload for the sensory, cognitive, and psychomotor workload components of each task identified
- develop computer-based workload prediction models using the data produced by the task analyses
- exercise the computer models to produce predictions of crew workload under varying automation and/or crew configurations

Each of the three phases in the refined methodology is described below:

Phase 1: Conduct Mission/Task Analysis

The first phase of the methodology is to conduct a comprehensive mission and task analysis for the proposed aircraft or simulator system. The mission/task analysis uses a top-down approach in which mission profiles for the system are subdivided into mission phases, and subsequently into mission segments. A segment is defined as a major sequence of events that has a definite start and end point. The events in a segment may occur concurrently or sequentially.

Each segment is then divided into functions. A function is defined as a set of activities that must be performed either by an operator or by equipment to complete a portion of the mission segment. Functions are categorized as continuous, discrete fixed, or discrete random and are placed on a rough time line using a Segment Summary Worksheet, such as the example selected from the AH-64A mission/task analysis (ref.2) and depicted in Figure 1.

The functions for each segment are subsequently divided into tasks. Each task is a specific crew activity that is essential to the successful performance of the function. The task consists of a verb and an object and is analyzed to

- identify the crewmember(s) performing the task
- identify the subsystem representing the primary man-machine interface
- estimate the workload imposed on the crew member(s)
- estimate the time required to complete the task

The crew member(s) performing each task and the subsystems associated with each task are identified by examining the manner in which similar tasks are performed in existing Army helicopters. Predictions of the visual, auditory, kinesthetic, cognitive, and psychomotor workload for each task are derived by writing short verbal descriptors of the requirements for each task component. The descriptors are then compared with the verbal anchors contained in the rating scales shown in the table (ref. 2). The rating (i.e., 1 - 7) associated with the anchor that best matches the verbal descriptor is assigned as the numerical estimate of workload. Two or more analysts perform the ratings independently and then reach consensus on the final ratings for each task. Task time estimates are assigned after interviews with subject matter experts (SMEs), or in some cases, after actual measurements of performance times on similar tasks.

Information derived from the mission/task and workload analyses is recorded on Function Analysis Worksheets, such as the one shown in Figure 2 for the AH-64A function "Fire Weapon, Missile" (ref. 2). The tasks are listed in the first two columns. The crew member performing each task is indicated by the letter (P for pilot; G for gunner; and B for both) that is presented in the third column along with a numerical identifier for the task. The subsystems associated with each task are presented in the fourth column. Verbal descriptors of the sensory, cognitive, and psychomotor components of workload and the ratings associated with each component are entered in the next three columns. The eighth column describes the type of switch for each task for which a specific switch is involved. The estimated length of time for discrete and continuous tasks is presented in the final two columns of the worksheet. The total time to perform all the tasks in the function appears in the upper right corner of the Function Analysis Worksheet.

Phase 2: Develop Computer-Based Workload Prediction Models

Phase 2 of the methodology consists of developing computer models to predict total workload experienced in the performance of both individual and concurrent tasks. The procedure used to develop the computer models represents a bottom-up approach in which the tasks identified in the Phase 1 mission/task analysis serve as the basic elements of analysis. Specifically, the information derived for each task is entered into computer data files from which estimates of total workload at the segment level are produced. Computer programs developed from time-based decision rules are then written to build functions from the tasks, and subsequently, to build segments from the functions. The decision rules define the temporal relationships among tasks and functions as determined in the mission/task analysis. By

implementing the decision rules, the computer models produce estimates of total workload, at half-second intervals, for each workload component (i.e., visual, auditory, kinesthetic, cognitive, and psychomotor). The estimates are derived by summing the ratings assigned to each workload component across concurrent tasks. A total value of "8" on any single half-second time line constitutes the threshold for an overload within a given workload component. A more detailed description of the Phase 2 methodology is provided in references 1, 2, and 3.

Phase 3: Exercise the Computer Models

During Phase 3, the computer models are exercised to predict workload associated with individual automation options and/or combinations of options. Three steps are performed to produce the workload predictions:

- select the automation options to be exercised by the model
- revise the estimates of workload for each task
- exercise the model to produce new workload predictions

The automation options are selected in consultation with engineers from the system program office responsible for acquiring the new aircraft or flight simulator. The tasks identified in the mission/task analysis are then reviewed to determine how each of the proposed automation options is likely to change the workload estimates in the baseline analysis. For each task affected by the automation options, new verbal descriptors of workload are written. These descriptors, in turn, provide the basis for assigning new workload ratings to the components of the tasks. New computer files containing the revised workload estimates are then established. Finally, the model is exercised with the new files to predict workload for any single automation option or any combination of automation options. Use of the model to predict crew workload for the LHX weapon system is described in detail in reference 1.

Application of the Workload Prediction Methodology

The methodology described above represents a systematic approach for predicting operator workload in advance of system design. As various automation options and alternative crew configurations are considered during the design of a weapon system, the methodology can be repeated so that the workload predictions keep pace with the system design process. Additionally, the methodology produces a number of

products that can be applied to the development of **any** complex weapon system. The products include:

- a mission/task/workload analysis that provides estimates of (a) sensory, cognitive, and psychomotor components of workload, and (b) performance times at the task level of specificity
- scales for rating sensory, cognitive, and psychomotor components of workload
- a timeline analysis that depicts concurrent crew tasks
- a procedure for evaluating total workload for concurrent crew tasks
- a numerical index for identifying crew overloads
- computer models that produce comparisons of workload for proposed alternatives in system design and crew composition
- a procedure for identifying an optimum design configuration for reducing crew workload

Workload predictions produced by the models have already been used by the Army in system trade-off analyses directed toward determining whether one or two aviators will be required to perform the LHX mission on the future battlefield and to assist in making decisions regarding the optimum configuration of LHX automation options.

VALIDATION OF THE WORKLOAD PREDICTION MODEL

The workload predictions yielded by the models have not been validated. Consequently, the next phase of the research will consist of (a) validation of the parameters used to develop the models, and (b) the validation of the workload predictions yielded by the models.

Parameters of the model that require validation include:

- workload ratings assigned to each task
- total workload estimates for concurrent tasks
- estimated times assigned to each task
- threshold for excessive workload
- temporal relationships among tasks
- procedural relationships among tasks

In designing the validation research a number of critical issues were considered. In this section, two of the critical issues most relevant to the workshop topic, *Mental-State Estimation,* are discussed and major provisions of the validation research plan are presented. A more complete discussion of the critical issues and a full description of current research plans are presented in reference 4.

Critical Issues

The problems and issues that have a critical bearing on the research required to validate the parameters in the workload prediction methodology include the following:

- reliability and validity of workload predictors
- selection of appropriate criterion measures.

Reliability and Validity of the Workload Predictors

The methodology used to derive the workload predictions requires that the reliability of both the rating scales and the predictors of workload be established. Specifically, it must be demonstrated (a) that the workload rating scales discriminate accurately between levels of attentional demand, and (b) that different raters will derive consistent estimates of workload for the sensory, cognitive, and psychomotor components of individual tasks. The reliability of the ratings assigned to the individual task components is important because these ratings are the basis for producing the predictors of total workload for concurrent tasks. If the individual workload ratings are found to have high reliability, the predictors of total workload produced by summming the ratings also will have high reliability.

The procedures used to develop the workload predictors are designed to ensure that the predictors have high face and content validity. The research for validating the workload model will attempt to establish that the predictors also have predictive validity. The predictive validity will be established by comparing the workload component ratings for each task, as well as the predictions of total workload associated with concurrent tasks, with (a) objective measures of primary task performance and (b) other subjective measures of workload. The primary task measures will be compared with the predictors at half-second intervals for each task on the mission segment timeline, while the subjective measures will be compared with the predictors for selected portions of the mission segments. Predictive validity will be demonstrated to the extent that the workload component ratings and/or the

total workload predictors correlate with the criterion measures.

Selection of Appropriate Criterion Measures

A number of performance measures will be selected as criteria for validating the workload predictors. Although evidence suggests that, in some instances, task performance may be relatively independent of workload (ref. 5), a critical assumption of the workload prediction model is that, when total attentional demand is driven close to or above the threshold of overload, performance on one or more of the concurrent tasks will be degraded. Consequently, the primary basis for selecting the performance measures to be used in the validation study will be their sensitivity to degradations in task performance due to increased workload. Additionally, the measures will be selected on the basis of their relevance to specific operator tasks. For example, deviations from a specified airspeed will be the criterion for workload encountered in the task "control airspeed." Such measures have high face, content, and construct validity.

Subjective measures of workload also will be collected during the validation research. The subjective measurements will be selected from among presently recognized and partially validated techniques, including (a) the NASA bipolar rating technique (ref. 6), (b) a modified Cooper-Harper rating technique (ref. 7), and (c) the subjective workload asessment technique (SWAT) (ref. 8).

Subjective measurements offer the system designer information that is not provided by the more objective techniques; furthermore, subjective methods of measurement are generally well received by operators and require little instrumentation. The greatest disadvantage of subjective workload measurements from the standpoint of the validation research is that the measurements do not provide information regarding the composition of the primary task. That is, it is just not feasible to collect subjective ratings at the task level of specificity. A second disadvantage is that subjective methods rely on the ability of operators to retrieve information from short-term and long-term memory regarding their experiences during task execution; yet, the behavioral literature is replete with examples demonstrating the fallibility of the memory retrieval processes (refs. 9 and 10). Even if the retrieval processes were reliable, it is not clear whether the recollections reflect task input modality (ref. 11), number of concurrent tasks (ref. 12), working memory load (ref. 13), or some other aspect of the task situation. Finally, empirical findings (ref. 14) suggest that retrospective subjective measures reflect the average workload experienced during task execution, thus precluding the analysis of workload at different points in time.

For several reasons there presently are no plans to employ physiological workload measurement techniques during the validation research. No single physiological measurement technique exists that is sensitive to task loading, diagnostic of task demand, **and** unobtrusive. A more serious problem with physiological measures is that they do not directly address the relationship between system design and workload, an important consideration on which system engineers base their design decisions. There are simply not enough data to establish whether the fluctuations of physiological measures actually reflect mental effort, some other operator "state" condition such as stress or fatigue, or a combination of several workload-related states.

The Validation Research Plan

The proposed research for validating the workload prediction methodology will be accomplished in three phases. During Phase 1, the reliability of the workload rating scales and the workload predictors will be evaluated. During Phase 2, validation data will be collected through a series of studies employing part-mission and full-mission simulation. During Phase 3, the results from Phases 1 and 2 will be used to refine the workload prediction model. Each of the three phases are described briefly below. More complete details are provided in reference 4.

Phase 1: Establish the Reliability of the Workload Rating Scales and the Workload Predictors

Phase 1 of the validation research will evaluate how closely the researchers' judgments in assigning numbers to the verbal anchors correspond with the judgments of other human factors scientists engaged in workload research. First, a psychophysical experiment using the method of paired comparisons (ref. 15) will be conducted by survey to (a) verify the ordinal ranks of the verbal anchors for each of the five workload component scales, and (b) produce equal interval scale values for each verbal anchor. Second, the empirically derived interval scale values will be applied to the workload component descriptors for all tasks. Finally, predictors of total workload will be produced by summing the interval scale values across concurrent tasks.

The human factors scientists also will be requested to rate the short descriptors of visual, auditory, kinesthetic, cognitive, and psychomotor components of workload for each task in the model. These same judges subsequently will be teamed in pairs. Each pair of judges will be instructed to assign a consensus rating for each of the verbal descriptors. Correlational techniques will be used to evaluate the

inter-rater reliability of the ratings produced by (a) each independent rater and (b) each pair of raters.

Phase 2: Conduct Part-Mission and Full-Mission Simulation

During Phase 2 of the validation research, both part-mission and full-mission simulation experiments are planned. The simulator configuration for both the part-mission and the full-mission simulation will be identical. For the part-mission simulation, mini-scenarios will be generated by selecting concurrent and sequential tasks from the mission/task analysis. An equal number of the mini-scenarios containing high- and low-workload sets of tasks will be selected. For the full-mission simulation, a composite mission scenario will be developed by selecting segments from the mission/task analysis.

The part-mission simulation will be conducted using a repeated measures experimental design in which each subject will fly the mini-scenarios multiple times. The order of presentation of the mini-scenarios will be counterbalanced to control for order effects and other extraneous variables. Analyses will then be performed to assess the correlation between the workload predictors and the performance measures recorded throughout the mini-scenarios. The correlation coefficients resulting from the analyses will serve as the primary measure of how accurately the workload predictors forecast excessive workload at the task level of specificity. Analyses also will be performed to assess the correlation between predictions of workload and subjective estimates of workload. These correlations will indicate the degree to which the workload prediction model predicts workload at the mini-scenario level of specificity.

To assess the validity of the time estimates used in the model, the actual amount of time required to perform the various tasks in the mini-scenarios will be compared with the estimated times produced during the task analysis. Differences will be resolved by adopting the recorded times. The time analysis will be used to validate the temporal relationships among the tasks as they exist in the workload prediction model. The procedural relationships among the tasks will be evaluated by noting the subjects' ability to progress through the mini-scenarios following the sequence of tasks specified by the model. Any new sequences adopted by the subjects to complete the mini-scenarios will be used to refine the workload prediction model.

During the full-mission simulation experiments, each trial will start at the beginning of the composite scenario and continue without interruption to the end. The analysis of results from the full-mission simulation will include all of the analyses performed during the part-mission simulation data analysis.

Phase 3: Refine the Workload Prediction Model

The final phase of the validation research will be to refine the workload prediction model. The first refinements will be made when the research results from Phase 1 are available. Additional refinements will be made when the part-mission simulation results are available; final refinements will be made when the full-mission simulation results are available.

CONCLUSIONS

Successful completion of the validation research will result in several useful products. The products will include (a) reliable and valid scales for predicting visual, auditory, kinesthetic, cognitive, and psychomotor workload at the task level of specificity, and (b) a validated workload prediction methodology that can be applied early in the system design process. Even without validation, the workload prediction methodogy proved useful during the trade-off analyses and other system studies conducted for the LHX. The baseline analyses currently being performed for the AH-64A, UH-60A, and CH-47 aircraft will benefit proposed modification programs for additional systems. After the validation research has been completed, the human factors community will have a tool with proven value for predicting operator workload early in the design of **any** proposed system.

REFERENCES

1. Aldrich, T. B., Szabo S. M., & Craddock, W.: A Computer Analysis of LHX Automation Options and Their Effect on Predicted Workload. Draft Technical Report ASI479-063-85. Fort Rucker, AL: Anacapa Sciences, Inc., August 1986.

2. Szabo, S. M. & Bierbaum, C. R.: A Comprehensive Task Analysis of the AH-64 Mission With Crew Workload Estimates and Preliminary Decision Rules for Developing an AH-64 Workload Prediction Model. Draft Technical Report ASI678-204-86(B), Vol I, II, III, IV. Fort Rucker, AL: Anacapa Sciences, Inc., October 1986.

3. Bierbaum, C. R.; Szabo, S. M.; and Aldrich, T. B.: A Comprehensive Task Analysis of the UH-60 Mission With Crew Workload Estimates and Preliminary Decision Rules for Developing a UH-60 Workload Prediction Model. Draft Technical Report ASI690-302-87(B), Vol I, II, III, IV. Fort Rucker, AL: Anacapa Sciences Inc., 1987.

4. Aldrich, T. B. & Szabo S. M.: Validation of the LHX One-Crewmember Workload Prediction Model. Draft Technical Memorandum ASI678-202-86(B). Fort Rucker, AL: Anacapa Sciences Inc., October 1986.

5. O'Donnell, R. and Eggemeier, F. T.: Conceptual Framework for the Development of Workload Metrics in Sustained Operations. Paper presented in Workload Assessment Methodology Workshop at the 27th Annual Meeting of the Human Factors Society, 1983.

6. Hart, S. G., Childress, M. E. & Hauser, J. R.: Individual Definitions of the Term "Workload." Proceedings of the Psychology in the DOD Symposium. Colorado Springs, CO: USAFA, 1982.

7. Sheridan, T. B., & Simpson, R. W.: Toward the Definition and Measurement of the Mental Workload of Transport Pilots. FTL R79-4. Cambridge, MA: MIT Flight Transportation Laboratory, 1979.

8. Reid, G. B., Shingledecker, C. A., & Eggemeier, F. T.: Application of Conjoint Measurement to Workload Scale Development. In R. C. Sugarman (Ed.), Proceedings of the 25th Annual Meeting of the Human Factors Society (pp. 522-526). Santa Monica, CA: Human Factors Society, 1981.

9. Crowder, R. G.: Principles of Learning and Memory. Hillsdale, NJ: Erlbaum Associates, 1976.

10. Klatzky, R. L.: Human Memory: Stuctures and Processes (2nd ed.). San Francisco, CA: W. H. Freeman, 1980.

11. Derrick, W. L.: The Relationship Between Processing Resource and Subjective Dimensions of Operator Workload. In R. C. Sugarman (Ed.), Proceedings of the 25th Annual Meeting of the Human Factors Society (pp. 532-536). Santa Monica, CA: Human Factors Society, 1981.

12. Wickens, C. D., & Yeh, Y.: The Dissociation Between Subjective Workload and Performance: A Multiple Resource Approach. In A. T. Pope and L. D. Haugh (Eds.), Proceedings of the 27th Annual Meeting of the Human Factors Society (pp. 244-247). Santa Monica, CA: Human Factors Society, 1983.

13. Gopher, D., Chillag, N., & Arzi, N.: The Psychophysics of Workload--A Second Look at the Relationship Between Subjective Measures and Performance. In R. W. Swezey (Ed.), Proceedings of the 29th Annual Meeting of the Human Factors Society (pp. 640-644). Santa Monica, CA: Human Factors Society, 1985.

14. Wickens, C. D., Yeh, Y., & Hart, S. G.: The Effect of Varying Task Difficulty on Subjective Workload. In R. W. Swezey (Ed.), Proceedings of the 29th Annual Meeting of the Human Factors Society (pp. 765-769). Santa Monica, CA: Human Factors Society, 1985.

15. Kling, J. W., & Riggs, L. A. (Eds.),: Experimental Psychology. New York: Holt, Rinehart, & Winston, 1971.

WORKLOAD COMPONENT SCALES

SCALE VALUE	DESCRIPTORS
	Cognitive
1	Automatic (Simple Association)
2	Sign/Signal Recognition
3	Alternative Selection
4	Encoding/Decoding, Recall
5	Evaluation/Judgment (Consider Single Aspect)
6	Evaluation/Judgment (Consider Several Aspects)
7	Estimation, Calculation, Conversion
	Visual
1	Visually Register/Detect (Detect Occurrence of Image)
2	Visually Inspect/Check (Discrete Inspection/Static Condition)
3	Visually Scan/Search/Monitor (Continuous/Serial Inspection, Multiple Conditions)
4	Visually Locate/Align (selective Orientation)
5	Visually Track/Follow (Maintain Orientation)
6	Visually Discriminate (Detect Visual Differences)
7	Visually Read (Symbol)
	Auditory
1	Orient to Sound (General Orientation/Attention)
2	Orient to Sound (Selective Orientation/Attention)
3	Detect/Register Sound (Detect Occurrence of Sound)
4	Verify Auditory Feedback (Detect Occurrence of Anticipated Sound)
5	Discriminate Sound Characteristics (Detect Auditory Differences)
6	Interpret Semantic Content (Speech)
7	Interpret Sound Patterns (Pulse Rates, etc.)
	Kinesthetic
1	Detect Preset Position/Status
2	Detect Movement (Discrete Actuation--Toggle, Trigger, Button)
3	Detect Movement (Discrete Adjustive--Rotary Switch)
4	Detect Movement (Continuous Adjustive/Flight Controls--Cyclic, Collective)
5	Detect Movement (Continuous Adjustive/Switches--Rotary Rheostat, Thumbwheel)
6	Detect Serial Movement (Keyboard Entries)
7	Detect Conflicting Cues
	Psychomotor
1	Discrete Actuation (Button, Toggle, Trigger)
2	Discrete Adjustive (Rotary, Vertical Thumbwheel, Lever Position
3	Speech
4	Continuous Adjustive (Flight Control, Sensor Control)
5	Manipulative
6	Symbolic Production (Writing)
7	Serial Discrete Manipulation (Keyboard Entries)

SEGMENT SUMMARY WORKSHEET

PHASE 3 Enroute

SEGMENT 08 Takeoff

PILOT			GUNNER		
DISCRETE (FIXED)	DISCRETE (RANDOM)	CONTINUOUS	DISCRETE (FIXED)	DISCRETE (RANDOM)	CONTINUOUS
Perform Hover (100) Perform Before Takeoff Check (091) Perform External Communication (099) Establish Climb (059) Establish Level of Flight (060)	Receive Communication (Internal) (116) Transmit Communication (Internal) (148)	Monitor Audio (078)	Perform Before Takeoff Check (090)	Receive Communication (Internal) (116) Transmit Communication (Internal) (148)	Monitor Audio (078)

Figure 1. Example of a Segment Summary Worksheet developed during the mission/task analysis (ref. 2).

FUNCTION ANALYSIS WORKSHEET

FUNCTION 065 Fire Weapon, Missile TOTAL TIME (Approximate) **5.5 Seconds**

TASKS				WORKLOAD COMPONENTS				DURATION (SECONDS) DISCRETE/ CONTINUOUS	
VERB	OBJECT	ID #	SUBSYSTEM(S)	SENSORY	COGNITIVE	PSYCHOMOTOR	SWITCH DESCRIPTION		
Verify	Firing Constraints	G239	Sensor Display (VSD)	Visually Discriminate Alignment Differences V-6	Evaluate Sensory Feedback and Verify Constraints Met C-2			1	
Pull	Weapons Trigger	B643	Weapons (AW)	Feel Trigger Movement K-2	Verify Correct Position (Trigger Activated) C-2	Lift Cover and Pull Trigger P-1	Springloaded Trigger (SPTR)	1	
Verify	Missile Launch	G417	Fire Control Computer/ Sensor Display (AFC/VSD)	Visually Detect Image V-1	Verify Correct Status (Missile Launched) C-2			1	
Release	Weapons Trigger	B644	Weapons (AW)	Feel Trigger Movement K-2	Verify Correct Position (Trigger Deactivated) C-2	Release Trigger P-1	Springloaded Trigger (SPTR)	.5	

Figure 2. Example of a Function Analysis Worksheet developed during the mission/task analysis (ref. 2).

OMIT TO
P. 99

INTRODUCTION TO SESSION II: STRESS AND STRESS EFFECTS

Robert P. Bateman
Boeing Military Airplane Company
Wichita, Kansas

During a discussion of some of the problems encountered in measuring mental workload, a psychologist recalled his first ride in an airplane. After a tour of the local area and a demonstration of the proper landing technique, his pilot asked if he would like to try a landing. Of course he would, and he did. However, on final approach to the runway, the pilot had to take control of the airplane and complete the landing. During a later discussion, the pilot asked if the psychologist had been worried during the landing approach. The reply was "no, of course not. Why?"

"Because we almost stalled out at a hundred feet in the air," was the reply. "That's why I took control." In this instance, there was a lack of stress in an individual because of a lack of knowledge of the situation. During his landing attempt, he felt that his stress and his level of workload were both low when, based on his performance and the danger involved, an outside observer might have surmised that he was highly stressed and working very hard to avoid a crash. In his own words, the psychologist was "too dumb to be afraid."

This anecdote provides a key to our definition of stress. An individual is not stressed because of the presence of stressors, but because the individual recognizes a situation in which there is a substantial imbalance between demands and the capability to deal with those demands. Since stress requires a recognition of the imbalance in the situation on the part of the individual, we consider it to be self imposed. When the stress on an individual produces measurable effects (external physical effects, internal physiological performance effects), we will call these effects strain. All stressors do not cause stress, and all stress does not cause strain. Some individuals can handle and/or tolerate stress much better than others.

In our investigations of mental states, it is not enough to assume that a reasonable person should be stressed. We need to know the degree of stress actually present in each subject in a specific situation. The knowledge of stress levels in individuals in real situations is essential in the design of aircraft crew stations.

In the laboratory, we use simulators to study pilot performance and mental workload. We would like to impose stress on the subjects because we want to use the data to project our findings to the real world. Nevertheless,we know that the acute stresses of flight are missing from our scenarios. There is a need for a transfer function which links experimental results to real world situations.

During our investigations of mental states, we have all experienced the loss of subjects who called in with headaches, personal problems, or stomach troubles. Regretfully, we have replaced subjects with these problems,

or adjusted our analyses to deal with the problem of unequal cell size. Since our subjects are usually volunteers, we have not been able to get the participation of people suffering with what we call severe chronic stress. Moreover, we have not developed a methodology for determining the state of chronic stress in our subjects who do participate. The first step in remedying this situation is to recognize the problem and the need for corrective action. It would be a wonderful world if all of our aviators were able to fly only when they were not stressed, but this is not the case. They fly with hangovers, headaches, and worries. They sometimes bring to the flight deck their financial problems, concerns for divorce, and grief over the loss of a loved one. Once a flight begins, these chronic stresses are joined by the acute stresses of aircraft emergencies, crowded airways, and information overload.

We do not pretend to have the answers for these problems. We do have some information to share on the nature of the problems of stress and its effects on the safety of flight. Our goal is to define the constructs we call chronic stress and acute stress and develop an increased awareness of the impact of these constructs on human performance and workload. In this session we will first discuss chronic stress. This will be followed by a presentation on acute stress.

N88-23376

S6-53
140794
7P.
NT 252649

CHRONIC STRESS AS A FACTOR IN AIRCRAFT MISHAPS

Robert A. Alkov
Naval Safety Center
NAS Norfolk, VA

Thirty four years ago, when I started Navy flight training, there was no such thing as pilot stress. The macho thing was that stress didn't bother you. Stress was for sissies. We were told by our flight surgeons that aviators compartmentalize their stress, keeping marital problems at home, office problems at the office, etc. In fact we were selected, in part, because of this talent. Unfortunately, this leads to the selection of individuals who are not very introspective, and not very aware of what's actually happening to them, physiologically.

In the late 1960s, I worked for a Navy flight surgeon at the Naval Aviation Safety Center, Captain Frank Austin. Frank recently resigned as Federal Air Surgeon. During the Vietnam war, in 1966, he took part in a study, (ref. 1) conducted by Drs. James Roman, Walt Jones and Harry Older of NASA, of the stress of combat on Navy pilots. They took a number of physiological measures of pilot stress - heart rate, respiration and so on. They actually instrumented aircraft so that they could tape these responses in flight. After a flight, they also did chemical analyses of blood, urine samples and so on. What they found was that while these individuals were over the target with SAMS (surface to air missiles) being fired at them, their stress levels went up pretty high as one would expect. But when they got back to the carrier just before landing, they went right off the scale on these measures of stress. There is no stress like landing on an aircraft carrier, especially at night - not even being shot at by the enemy. But, the pilots were able to handle this type of acute stress better than the average person.

Naval aviators pride themselves on being at the "tip of the spear," of U. S. policy. Unlike the other services, they have to be ready and in place, near the battle zone, with their aircraft loaded with ammunition ready for trouble at any time. This means that they're in a constant state of readiness. They don't really feel they have time to talk about things like stress.

When you start looking at what these Navy personnel are doing, you can't really question that they are under conditions that produce chronic stress because we are talking about long family separations and severe living conditions. If any of you have ever been aboard an aircraft carrier you know what I am talking about. You are working in an environment of noise, temperature extremes, and vibration. There's really no rest. As for sleep, in many cases during the Vietnam war, "hot bunking" was practiced. Since there are accommodations for a fixed number of sailors, but more people than that were required to do the job, bunk sharing had to be employed. When a sailor was on duty and out of the bunk, his buddy was sleeping in it. He really had no place to go where he could relax. On board an aircraft carrier you are working, or you are sleeping, or you are eating a meal. Usually the meals are on-the-fly affairs, 15 minutes at the most. There's not a whole lot of rest. Work days are often 18 hours or

more. These conditions are conducive to chronic fatigue and stress in maintenance and aircraft handling deck crews.

A naval aviator's main duty is not just flying. He also has collateral duties. There is limited space aboard ship to put personnel, so every officer has to double up on jobs. In addition to being a pilot he also may be a maintenance officer, or an operations officer, or a training officer, or a safety officer. He may get so bogged down in his paper work that there's just no time to think about flying or to study flight manuals while at sea. There's a lot of uncertainty in the kinds of operations that may be assigned when a carrier air group is tasked to react to some kind of external threat. They may be at the point of returning home from a long sea period, when they're suddenly turned around and sent back. This seriously affects their family life. They find themselves involved in "blue water operations". That is, they are so far from land that, if anything happens, they can't go back and land somewhere on a land base. (At least not on friendly territory where you have prior arrangements to land.) So if anything happens and they can't get aboard the aircraft carrier they have to ditch at sea or eject from their jets. Many naval aircraft mishaps involve aircraft that take off and are never heard from again. Unfortunately, these mishaps can't be investigated to determine causes and correct them.

Also, there exists the potential for chronic stress that is caused by the heavy responsibility that's put on junior officers. The Navy utilizes independent duty detachments with fairly junior officers, especially in the helicopter community. They might be placed aboard a ship at sea that is not an aviation ship. It just has a very small platform built on the stern. Many times the commanding officers of these ships have no aviation background. They may want a crew to fly over and pick up some vitally needed supplies or take a wounded or sick man back to shore. It'll be a dark night, in freezing weather, and the deck will be rolling and pitching. A junior officer has a lot of pressure put on him to complete these kinds of missions. The same thing happens to Coast Guard aviators who are tasked with missions to rescue people under similar conditions.

The rotary wing community has been largely neglected when it comes to safety. Historically, the emphasis in human engineering design for safety has been on the more expensive fighter and attack aircraft. The individuals who are out there in small helicopter detachments, hovering over a very slippery deck at night in rough weather, have been overlooked.

But naval aviators are stress copers. They thrive on it. They're selected for this. In fact, they are stress seekers. You could call them type A personalities, people who have to be under pressure to really do good work. However, each has his own personal limitations. Stress coping is subject to individual differences. If you drew a curve showing their stress coping behavior, it would be a bell shaped curve. But, that curve would represent higher overall stress coping ability than that of the normal population. We may think that an individual aviator is doing well compared to the general population but we would have to compare him to the norm for his group to say he is coping well. Those who do not cope well represent a small percentage of our aviators, however.

We find that people who do not cope well with stress tend to fall into two categories. First there is the younger and the less experienced, immature individual as you would expect. These represent a substantial portion of the sailors who man the flight decks and do maintenance on the aircraft. Secondly, the type A personality frequently has trouble coping with stress. This description would fit a lot of our junior aviators. These two groups do not handle stress well.

I got into stress coping research because there was not widespread recognition in the Navy of stress as a mishap cause factor. Even Navy flight surgeons who are trained to do the human factors analysis on aircraft accidents were not recognizing stress when it was a factor in a mishap. Several years ago this was demonstrated during the investigation of an accident that involved an aircraft commander taking off in a transport aircraft who had an engine quit. The copilot, using good crew coordination procedures, tried to feather the bad engine. The aircraft commander reached up and knocked his copilot's hand away from the engine feather button, then proceeded to feather the good engine. They ditched into the sea and got out alive, but they lost a couple of passengers who drowned. The flight surgeon had written up his report declaring that there were no known psychological or sociological factors in the mishap. The Naval Safety Center's aircraft accident investigation team was sent to investigate. One of the team was in the officer's club bar at the base where the accident occurred. He started asking some questions and found out that the pilot was in the bar the night of the mishap. He finally had to leave when the bar closed at one a.m. He had an early flight at six a.m. and he had been drinking heavily. The reason for this was his wife had called him from the United States to tell him she was leaving with another man. This was the culmination of months of marital discord. Apparently everybody knew about it in his squadron. They just closed ranks and were tight-lipped about it during the investigation, to protect their buddy.

About that time I began talking with Captain Richard Rahe, a psychiatrist at the Naval Health Research Center in San Diego. He's now retired and teaching at the University of Nevada Medical School in Reno. I asked him if the life changes scale that he had determined was associated with health changes could also have some correlation with behavioral changes. Some of these health changes include accidental injuries. Certainly if life changes had such a profound effect on health there surely must be some effect on skilled performance. However, since his interest was only on health changes, Dr. Rahe encouraged me to investigate a relationship, if any, between life events and accidents.

You probably recall the study in which Dr. Rahe collaborated with Dr. Thomas Holmes of the University of Seattle. They had a large number of faculty members rank-order various life changes as to how much stress coping they felt would be required by each. They arbitrarily assigned the death of a spouse (the one that everyone agreed required the most stress coping) 100 points using an ordinal scale. Thus divorce was assigned 73 points and so on. In the Navy study, Captain Rahe added the cumulative points for people who reported these kinds of events within a year prior to going on a cruise on U.S. Navy ships from San Diego. There were over 2,000 men involved in this study. They weren't told why they were being asked these questions. During their cruise, of those who had accumulated between 150 to 200 points,

about a third reported to the sick bay with some kind of illness. If they had between 200 and 300 points, over half reported ill during the cruise. Of those with over 300 points almost 80% reported ill or with some kind of accidental injury (ref. 2).

Even though ordinal scales are not additive, they did demonstrate a relationship between cumulative life events and health changes. So I devised a questionnaire of my own. I found that a lot of these life change factors didn't work for me in discriminating aviators who had pilot factor mishaps (ref. 3). Again, there were too many individual differences, so I started looking beyond life changes to such things as stress coping. The questionnaire I used asked about pilot judgment and life difficulties as well as certain personality characteristics (ref. 4).

My questionnaire was adapted from Drs. Rahe and Holmes's list of life events, plus some biographical information and data on aviator performance. It was sent to flight surgeons who were on aircraft mishap boards. They were instructed to complete the questionnaire for the involved aviators. By talking to his family and friends, his superiors in the squadron, his peers, etc., the flight surgeon could get the answers without showing the aviator the questionnaire. Many times the pilot was deceased, so the information had to be obtained from the family. The pilot never saw the questionnaire, only the flight surgeon did. Unfortunately, it was an ex-post-facto study. This has led to a great deal of criticism of the study. At the time a mishap occurs we don't always know exactly what happened so we have an investigation. By the time the causal factors are determined, a year might have elapsed. But when the investigation was finished the questionnaire responses were divided into two groups. Over 700 of these questionnaires were completed. They were roughly divided into half between those with a pilot error factor assigned and those who had no role to play in the cause of the mishap. (Roughly half of major Navy aircraft accidents are determined to be caused by pilot error.) Those that were assigned pilot error by the aircraft mishap boards were compared with those who had no fault in the mishap.

The results are shown in table 1. Several of the factors are related to having problems with interpersonal relationships, i. e., having problems with peers, problems with superiors, etc. (ref. 5). I have recently collected data from people who have not been involved in a mishap by asking the flight surgeon to use the same questionnaire on an individual in the squadron who's the same rank and roughly the same experience as the mishap-involved aviator. I have not published these results yet, but I can say that those people who have not been involved in a mishap are not statistically different from the group that were not at fault in the mishaps they were involved in. Both of these groups are statistically different from the at fault group in certain areas in the same direction as the previous studies.

The study identified some of the symptoms of inadequate stress coping that are associated with a pilot factor aircraft mishap. These include difficulties with interpersonal relationships (i.e. peer troubles and problems with authority figures). The mishap itself is also a symptom of inadequate stress coping. When an individual is not coping, he may turn his frustrations inward and become self destructive or he may "act out", taking out his feelings on others or on objects around him. The aggressive

personality characteristics exhibited by most aviators lend themselves to "acting out". My results demonstrated that "acting out" behavior was present in the at fault mishap pilots at the time of the mishap (ref. 5).

Sloan and Cooper in Great Britain attempted a study to determine if my results were applicable to British airline pilots (ref. 6). They sent my questionnaire to mature airline captains (average age in their late forties) and asked them which characteristics they thought would be important in identifying accident prone pilots. "Acting out" symptoms were not among them. Since their methodology was so completely unrelated to mine, I find the results are not comparable. My subject population consisted of young (average age 29 years) aviators who had been involved in aircraft mishaps. They never saw a questionnaire. The data were collected by flight surgeons trained to investigate by asking questions of supervisors, family, friends and fellow aviators. Sloan and Cooper's subjects, on the other hand, were asked to make a subjective evaluation.

What would I recommend? I think that in spite of the fact that this study was of military aviators, a highly select group, there are still a lot of lessons to be learned for general and commercial aviation. I'd like to list for you some characteristics of what I feel are successful stress copers. These people have a higher degree of self-awareness and feelings of self-worth. They believe they can influence events or even change them. Change is seen as a challenge and an opportunity, rather than a threat. As for recommendations for coping with stress, I believe the traditional methods aviators employ, which usually involve alcohol consumption, are counterproductive, causing more problems than they alleviate. Instead rest, exercise and a proper diet should be encouraged. In other words, physical fitness is a better strategy for stress coping.

Also recommended is time management, the prioritization of life goals, more self awareness, stress avoidance and counseling by a flight surgeon or chaplain if needed. The idea that only sissies are affected by stress must be put to rest. The subtle and insidious effects of stress on pilot performance must be emphasized in pilot training programs. Stress and fatigue are hazards that must be dealt with in aviation to ensure safe flight operations.

REFERENCES

1. Roman, J., Older, H., and W. L. Jones, Jr.: Medical Monitoring of Navy Carrier Pilots in Combat. Aviation, Space, and Environmental Medicine, vol. 38, February, 1967, pp. 133-139.

2. Rahe, R. H.: Life Crisis and Health Change. Naval Medical Neuropsychiatric Research Unit Report No. 67-4, 1967.

3. Alkov, R. A., and M. S. Borowsky: A Questionnnaire Study of Psychological Background Factors in U.S. Naval Aircraft Accidents. Aviation, Space, and Environmental Medicine, vol. 51, September, 1980, pp. 860-863.

4. Alkov, R. A., Borowsky, M. S., and J. A. Gaynor: Stress Coping and the U. S. Navy Aircrew Factor Mishap. Aviation, Space, and Environmental Medicine, vol. 53, November, 1982, pp.1112-1115.

5. Alkov, R. A., Gaynor, J. A., and M. S. Borowsky: Pilot Error as a Symptom of Inadequate Stress Coping. Aviation, Space, and Environmental Medicine, vol. 56, March, 1985, pp. 244-247.

6. Sloan, S. J., and C. L. Cooper: The Impact of Life Events on Pilots: An Extension of Alkov's Approach. Aviation, Space, and Environmental Medicine, vol. 56, October, 1985, pp. 1000-1003.

Table 1. Factors which discriminated between pilots who were at fault in an aircraft mishap and those who were not at fault using the Fisher-Irwin Exact Test (one-tailed). (N=737)

Factor	At Fault (n=381)	Not at fault (n=356)	Critical Level (1 sided)
Poor leader	43	21	0.0065**
Lacks maturity and stability	20	9	0.0425*
Financial problems	14	5	0.0418*
Recent marital engagement	17	5	0.0118*
Recent major career decision	77	36	0.0001**
Difficulty with inter-personal relationships	26	13	0.03858*
Trouble with superiors	27	5	0.0001**
Trouble with peers	19	7	0.0203*
Recent personality change	13	4	0.0304
Excessive alcohol use or recently changed intake	8	0	0.0047**
No sense of own limitations	26	11	0.0131*
Incapable of quickly assessing potential troublesome situations	31	6	0.0000**

*p<0.05 **p<0.01

N88-23377 S7-53 140795 4P BR805365

ACUTE STRESS

Robert P. Bateman
Boeing Military Airplane Company
Wichita, Kansas

Acute stress may be defined as the realization that the immediate environmental situation has placed demands on an individual which possibly will not be handled successfully. It differs from chronic stress in that the demands and the effects of a failure to handle the demands are localized in time. There are optimal levels of acute stress which cause an optimal level of arousal and performance. When the level of acute stress exceeds these levels, accidents occur.

Consider the following hypothetical accidents:

A student pilot in a T-38 starts his engines for a solo flight, but an alert crew chief notices that the gear doors do not close. When he cannot communicate the problem to the student, the crew chief has the student shut down the left engine while the crew chief climbs up to the cockpit to reset the proper hydraulic switch. Shutting down the engine causes the stability augmentation system to drop off line. The student then restarts the left engine and taxis to the runway. He does not reaccomplish the check list, and takes off with the stability augmentation system off. During the climb, he notices the lack of a stability augmentation system. He levels off without reducing power and attempts to engage the stability augmentation system. In doing so he enters the regime of flight for maximum sensitivity for the flight controls and a small disturbance produces a vertical pilot induced oscillation. The aircraft goes through a violent maneuver during which the bearings on one engine fail. The student manages to recover by letting go of the stick, and decides (correctly) to shut down the failed engine and return to base.

During the violent maneuver, the seat cushion has been raised out of the seat pan, and is now lodged over the forward lip of the pan, tilting the cushion back and preventing the student from reaching the rudder pedals. Under the acute stress of the situation, the student fails to appreciate the implications of this situation. Without being able to reach the rudder pedals he will be unable to correct for the yaw during single engine operation. Upon landing he will be unable to steer the aircraft or to apply the brakes to stop it.

Under stress, his mental processes narrow down to one factor: Get this plane back on the ground! He does not remember that the gear will have to be lowered by the alternate system because he has shut down the engine, which normally provides hydraulic power for gear extension. In this state, he also forgets to calculate the correct airspeeds for base turn and final approach, corrected for his heavy fuel state. Nevertheless, when he contacts the tower to advise them of the situation, he states that, other than the loss of engine, he has no problem.

He begins his base turn two miles from the runway, thirty knots below the correct airspeed, pulling hard on the stick to keep his altitude. The only thing that keeps him from stalling the aircraft is the hard seat cushion draped over the seat pan which prevents him from getting the stick full back.

Realizing the aircraft is slow, he tries to light the afterburner on the good engine. This is the only way to save the aircraft, but at the low airspeed, a high power setting causes the aircraft to yaw and lose lift on one wing. Unable to correct the yaw or the resulting roll, the student takes the engine out of after-burner.

The aircraft struck the ground at a high rate of descent, about one mile from the end of the runway. After surviving this highly stressful accident, the student stated that the touchdown was mild, but that the gear may have failed during the landing roll because he remembered that he didn't have to climb down very far to exit the cockpit. Actually, the gear sheared off at ground impact and was thrown over half a mile from the impact site. The aircraft came to a stop when it plowed into a sand dune, and with sand pouring over the canopy rail, the student would have had to climb UP to exit the plane. An excessive level of acute stress can impair mental processes. In this situation, it is reasonable to conclude that the student was stressed, despite his apparent lack of concern for the seriousness of the situation. Stressors do cause stress.

Consider a test pilot who stretches a mission in order to accomplish the planned maneuvers for a test program which is behind schedule. Enroute to his recovery base, the last fuel tank fails to feed into the main tank. The failure is not the pilot's fault, but it was his decision to continue the test below planned fuel minimums which produced the critical situation when the failure occurred. He declares an emergency and heads for the nearest usable runway. He is the recognized expert on this aircraft, and was the pilot who developed and tested the precautionary flame out pattern. Upset with himself because he will not recover as planned, he misreads his airspeed by 100 knots, decides not to go around because of his low fuel state, and runs off the far end of the runway. The situation should have been routine for this pilot. It should not have caused an excessive amount of stress, but it did. The stress was self imposed.

Consider an instructor pilot with previous experience in an aircraft which had a tendency to blow up when the engine caught fire. Suppose he lost a few close friends because they were a bit slow in deciding to eject. Now, put this pilot in a new aircraft with a student. Give him a fire light. Let him roll into a tight turn and ask the student if he sees smoke trailing behind the aircraft. Introduce a student who is not familiar with the condensation that forms in the wingtip vortex under a high g turn. The student reports the "smoke" that he was instructed to see. Student and instructor both eject safely. The airplane crashes - unnecessarily. The stress on the instructor was acute and severe. It may also have contributed to the student's error. A fire light in the trainer aircraft was not supposed to cause that much stress. The fact that it did underlines the previous conclusion: stress is self imposed.

Consider a student pilot in a T-37 on an initial solo flight away from the traffic pattern. During a practice approach to a stall, the left wing drops. He becomes preoccupied with raising the wing, using both aileron and rudder. This causes the approach to a stall to become a full stall and enter a spin to the right. The entry to a spin from a level attitude can be disorienting and frightening, especially when it is unintentional. Since the last recognizable attitude was left wing low, it should not be surprising

to learn that this stressed individual thought that he had entered a spin to the left. Since we have learned that stress tends to impair the mental process, it is understandable that the presence of a ball that is fully deflected to the left was interpreted as evidence confirming the "left" spin. In the T-37 aircraft, the ball on the left turn and slip indicator always moves left in any spin. We should not be surprised to learn that the airplane crashed. Amazingly, the student pilot did eject in time to survive.

Not all pilots survive stressful situations. Consider the pilot of Blue Four, the fourth aircraft in a formation scheduled to practice a bombing attack during marginal weather. The planned activity was radar bombing, which can be accomplished quite well in poor weather. When in actual combat situations, radar bombing is done one ship at a time; during peacetime, we fly to a range in formation. It's not supposed to be more hazardous in formation; it just turned out that way. In addition to watching his radar scope inside the cockpit, number four in a formation must also watch out for the other aircraft in formation. One more thing: in the interest of range safety, the pilots on this flight were required to acquire the target visually to insure that they did not drop bombs on the wrong target. In combat, that's not required. Radar bombing of unseen targets at night and in weather is often accomplished.

Blue Four proceeds to the target area, watching outside for the rest of the formation, watching his radar scope for the target, and trying to look outside to visually identify the target. The weather is marginal. The flight leader breaks up the formation on the range and begins his run. He has a good radar return, but cannot identify the target visually. Blue Two and Three make their runs with the same results. The leader decides to abort the bombing mission and calls for the flight to rejoin. Two and Three manage to find their leader, but Blue Four does not answer. He has crashed, wings level, slightly nose low, looking for the target - or for the rest of the formation. We'll never know. Blue flight returned without him.

We are reasonably sure that Blue Four was overstressed because we can talk with the other members of the flight, and they were stressed. We do not know how many of this year's accidents occurred because of stress. We can't always talk with the people involved. However, we are reasonably sure that acute stress is a factor in aircraft accidents.

Aviation will always provide a stressful environment. We probably can't change that, but before a stressful environment can result in an accident, the human must impose a level of stress upon himself, and then fail to handle the combination of the situation and the stress, resulting in a measurable strain: a failure in the system.

Against this background, there are some questions to consider:

- Can we identify an individual who is under stress?
- Can we keep a person with chronic stress away from stressful situations?
- Are chronic and acute stress additive in nature?
- Can we determine how much stress a specific individual can tolerate before a strain results?
- Can we reduce the stress before a strain results?

There may not be answers to these questions at present. They serve to define a problem. If we can agree on the problem and can understand the importance of solving the problem, then we have taken the first step towards its solution.

N88-23378

PUPIL MEASURES OF ALERTNESS AND MENTAL LOAD

Richard W. Backs and Larry C. Walrath
Human Performance Laboratory
McDonnell Douglas Astronautics Company
St. Louis, Missouri

1.0 OVERVIEW

As part of our internal research and development program at McDonnell Douglas we are examining human factors engineering issues associated with how operators extract information from visual displays. Recently, we have been using psychophysiological measures of operator performance, in addition to behavioral performance measures, in order to better assess operator mental workload (MWL) associated with using a particular display configuration during performance of a visual search task.

In our work, we take a rather broad view of the concept of MWL. That is, we consider MWL to be the cognitive effort associated with performing an information processing task analogous to the physical effort required to perform a manual task. The problem with such a definition, of course, is specifying precisely what is meant by "cognitive effort." We assume that cognitive effort is determined by the extent to which the information processing resources required to perform the criterion task are actively engaged in task performance. This definition presupposes that the task can be performed within the limitations of the available resources. Unfortunately, in practice MWL very often becomes synonymous with the paradigms with which it is manipulated (such as primary and secondary tasks) or the dependent variables with which it is measured (such as behavioral performance decrements in reaction time and error rate).

Clearly, there are many determinants of MWL. Two of these are the nature of the task and the required behavioral performance. Another determinant is the capability of individual operators to allocate their processing resources in ways to efficiently and effectively perform the task. This ability to optimally allocate resources requires a combination of the operator's natural abilities, training, and motivation. Any time there is a mismatch between the optimum level of resource allocation required by the task and the optimum level at which the operator is able to engage the necessary resources, an unacceptable amount of MWL will result. This mismatch may occur either because the task requires too much or too little cognitive effort.

Further, MWL is a closed-loop process, and as such is also determined by the costs to the operator (in physiological terms) of maintaining performance. These costs are increased in tasks that require either more or less than the operator's optimal level of cognitive effort. The physiological

costs become part of a feedback loop, along with the knowledge of results of the behavioral performance, and they serve as additional inputs upon the operator.

The MWL itself is, just as clearly, not a unitary phenomenon. Inappropriate load on the operator may occur at any of a number of points in the information processing flow. Although we are not testing psychological theory, we make heuristic use of several theoretical models. The first is that total information processing capacity is divided into multiple resource pools according to sensory input channels (e.g., ref. 1). The second is that information processing occurs serially, progressing through well-defined stages that can be manipulated independently (ref. 2). We further assume that, with the exception of those stages requiring access to common resources that must be shared, information processing can progress independently within each resource pool and in parallel with similar ongoing stages in other resource pools (ref. 3). We recognize that overall task performance is determined by the number and priority of sensory input channels required by a task and the amount common resource time-sharing required for task completion. However, up to this point in our research, we have not been concerned with concurrently manipulating multiple sensory channel resource pools or with the competition between pools for common resources.

We believe that in order to accurately assess an operator's MWL it is necessary to measure as many of its facets as possible. Monitoring behavioral performance is absolutely necessary since this measure is the end product to be maximized. Subjective reports of MWL can be helpful to define which elements of a task operators have trouble with. Subjective reports may also indicate circumstances in which objective measures fail to reflect deficiencies in workload and thus more sensitive objective measures are required. In our research, we use psychophysiological measures to provide such a sensitive measure. An added benefit is that the psychophysiological measures serve as a window into how the operator is allocating resources. Our goal is to discover which external (task) determinants contribute to MWL and which internal (cognitive) processes are inappropriately loaded.

As an example of our progress toward assessing operator MWL during visual search, we will present data from a recent study measuring evoked pupillary responses and response time to search displays that varied with regard to their density, use of color coding, and type of information abstraction required to complete the search. This study consisted of a single task, and was one of a series of studies originally designed to evaluate the effects of different display parameters on search time. It is meant to serve as an illustration of how adding psychophysiological response measures can help localize points of mental overload.

In a previous study (ref. 4), we described how eye-movement analysis was used to determine the effects of information density, use of color coding, and type of information abstraction on visual display search time. In that study, we found that search time and the number of fixations required to search a display increased with the density of the display. Longer search times and more fixations were also required to count the number of target items in a display than to locate a single target. However, even though

search time was longer for monochrome than for color-coded displays, the number of fixations required to search these displays did not differ. Instead, the duration of each fixation was shorter for color-coded than for monochrome displays indicating that subjects processed symbolic information more efficiently using a color code than using a shape code.

We also obtained evoked pupillary responses in reference 4 in order to evaluate this measure as an indicator of information processing load (e.g., refs. 5 and 6). Single-trial pupillary responses observed in reference 4 had a distinctive tri-phasic shape (dilation-constriction-dilation) similar to the average pupillary response data reported in reference 7. Significant effects of color coding and color coding by type of information abstraction were obtained for the initial dilation-constriction phase following display onset. However, an uncontrolled change in luminance preceding the search display was subsequently discovered. That change could possibly have accounted for the unexpectedly large constriction. In the present study, the luminance problem was corrected and the basic search task was repeated on another sample of subjects. In addition, these subjects participated in a psuedo-search condition which was included as a control for nontask-related luminance and color effects of the displays.

2.0 METHOD

2.1 Subjects

Eight McDonnell Douglas Corp. employees participated as subjects. Two of the subjects were female, and the age of all subjects ranged from 19 to 42 years. One subject had participated in reference 4, and another subject had previously completed the search task; both of these subjects were placed in the group that received the active search condition first. All other subjects were naive to the experimental procedure.

2.2 Apparatus

A Data General Eclipse S-140 minicomputer was used to generate the stimulus displays, control and time the experimental events, and collect and reduce for analysis the pupil diameter and response time data. Displays were presented on an AED 512 high-resolution color graphics terminal. Pupil diameter data were collected at 60 Hz using an Applied Science Eye View Monitor and TV Pupillometer System model 1994-S. The experimental set-up is shown in Figure 1. All photometry to calibrate luminance of the stimulus displays was performed with a Photo Research Co. Spectra-Pritchard Model 1980-A photometer using a photopic filter.

2.3 Procedure

Subjects participated in two experimental sessions: an active search task (SEARCH) where they were required to abstract information from a display, and a passive psuedo-search task (CONTROL) where they received the same task as in the SEARCH condition but were not required to abstract information from a display. SEARCH and CONTROL conditions were administered on successive days. Half of the subjects (one female) received the SEARCH condition first, while the other half received the CONTROL condition first.

Subjects viewed four different displays for each combination of the Information Density (10 vs 20 symbols), Color Coding (redundant with symbol shape vs monochromatic symbology), and Search Type (COUNT vs LOCATE a specific symbol: requiring exhaustive or self-termininating search strategies, respectively) independent variables for a total of 32 trials in both the SEARCH and CONTROL conditions. The order of presentation for the 32 displays was determined randomly for each subject in both experimental sessions.

Trials consisted of a series of four screens. The first was a calibration screen with a central fixation point and four calibration points that defined the 8.8° square area of the display containing the symbology. The second was a question screen, presented for 6 sec, identifying the search type and, in the SEARCH condition, the target symbol. The target symbol was always presented in the color in which it would appear in the display (i.e., yellow rectangles, red triangles, and green semicircles for the color-coded condition or all green symbology for the noncoded condition). The third screen was the calibration screen. The display screen was presented only if subjects fixated within 1° of the central fixation point for 0.5 sec during the calibration screen. If no such fixation occurred within 2 sec, the question screen was presented again and the trial was repeated until the subject did fixate on the central point. The fourth screen was the display, which was presented only after central fixation had been verified. Figure 2 contains examples of question, calibration, and high and low density display screens.

The procedure in SEARCH and CONTROL conditions was identical except for the search and response instructions given to the subject. In the SEARCH condition, subjects actively searched the display for the target and made a button press, which terminated the display, to indicate that they had completed their search. This response time to search the display was measured in msec from display onset. Subjects then verbally reported the number of targets (for the COUNT trials) or the the quadrant of the display in which the target was located (for the LOCATE trials). Whenever subjects failed to complete a search within 6 sec, the display screen was replaced by the calibration screen and they were required to guess at the correct answer. In the CONTROL condition, subjects were not given a target to search for on the question screen; instead, they were told to merely scan each display until it terminated. Also, subjects had no responses to perform. The experimenter controlled the length of the display screen, varying it from 2-6 sec, and no verbal response was necessary.

The 32 different display screens were approximately balanced with respect to the distribution of symbols, the location of targets within the four quadrants, and the frequency of the correct answer (1, 2, 3, or 4 targets in the COUNT condition and quadrants 1-4 in the LOCATE condition). Luminance of all text and symbology on the displays was equated at 0.51 fL. Overall screen luminance within the 8.8° search area was equated for all screens (at 0.52 fL) by varying background luminance. Ambient illumination was 8.49 x 10^{-2} ft-c.

2.4 Data Quantification

Single-trial pupillary responses exhibited the characteristic tri-phasic shape previously reported (refs. 4 and 7). Figure 3 shows representative

single-trial responses from a low density, color-coded trial and a high density, noncoded trial. Several measurements were made for each trial, baseline (pupil diameter at display onset) and three "components" (points of inflection for dilation or constriction). The first component (D1) was a small initial dilation that peaked about 266 msec after display onset. The second component (C) was a large constriction that peaked about 941 msec after display onset. These components were followed by a gradual dilation (D2), the resolution of which depended upon display duration. The differences between the D1 and C components and the D2 and C components were also computed for analysis. The D1-C difference was computed to determine the relative size of the constriction from the point of onset. The D2-C difference was computed to determine the amount of pupillary dilation that occurred from the point of maximum constriction. If the point of maximum dilation did not occur prior to the motor response, then the last data point in the trial was used as D2. Each of these measures and the search time were averaged over the four trials of each combination of Information Density, Color Coding, and Search Type.

All analyses were performed with the SAS General Linear Models procedure (ref. 8). A Latin square (ref. 9) was used to balance the effects of Group (SEARCH or CONTROL condition first), Condition (SEARCH or CONTROL), and Day (first or second test day), while the effects of Density, Color Coding, and Search Type were totally within-subjects. The degrees of freedom for all F ratios were (1,6) with the comparison-wise error rate set at $p < 0.05$. Duncan's Multiple Range tests were performed for all significant main effects and two-way interactions using the SAS Duncan procedure.

3.0 RESULTS

The main effect of Condition ($F = 11.52$) was significant for the baseline measure, reflecting the overall larger pupil diameter in the SEARCH than in the CONTROL condition. This effect was probably due to a generalized arousal difference between the two conditions as it was significant for all component measures. In order to correct for this initial difference, the baseline was subtracted from each component prior to analysis. Where results for component and peak-to-peak difference scores overlap, we will report only the peak-to-peak data.

The peak-to-peak difference scores, D1-C and D2-C, were both affected by the Condition and Color Coding manipulations, but in distinctly different ways. As shown in Figure 4 (left panel), the main effects of Condition ($F = 13.28$) and Color Coding ($F = 88.83$) were significant for the D1-C component, and these effects did not interact. Pupil diameter was larger overall (i.e., the size of the constriction was smaller) in the SEARCH than in the CONTROL condition, and pupil diameter was also larger for noncoded as opposed to color-coded displays. However, for D2-C (Figure 4, right panel), only the Condition by Color Coding interaction was significant ($F = 11.30$). Although none of the pair-wise comparisons differed significantly, pupil diameter for the D2-C component was larger for noncoded than for color-coded displays in the SEARCH condition, consistent with the D1-C data. However, in the CONTROL condition, pupil diameter was larger for the color-coded than for the noncoded displays.

The Condition by Search Type interaction was significant for both D1-C and D2-C (F = 9.14 and 18.37, respectively). The form of the interaction, however, was quite different for the two components. For the D1-C component (Figure 5, left panel), pupil diameter was larger (i.e., less constriction) in the SEARCH than in the CONTROL condition, and the difference between SEARCH and CONTROL conditions was greater in the LOCATE (self-terminating search) than in the COUNT (exhaustive search) trials. For the D2-C component (Figure 5, right panel), there was a crossover interaction in which no comparisons between means differed significantly. However, pupil diameter in the SEARCH condition was larger (i.e., greater dilation) in the COUNT than in the LOCATE trials.

The interaction between Density by Color Coding was significant for the D2 component (F = 11.09). As can be seen in Figure 6, pupil diameter for color-coded displays was larger for high-density than low-density displays. The opposite was found for noncoded displays, with larger pupil diameters found for the low-density displays. The difference between high- and low-density displays was not significant in either color-coding condition, however.

Search times (from the SEARCH condition) were significantly shorter for low vs high density displays (F = 42.52), for color-coded vs noncoded displays (F = 34.08), and for LOCATE vs COUNT trials (F = 16.18). However, the Density by Search Type (F = 10.52) and Color Coding by Search Type (F = 16.54) interactions were also significant. Search times were faster for low than for high density displays for both COUNT and LOCATE trials, but this difference was much greater for COUNT trials. Similarly, color coding decreased search time for both COUNT and LOCATE trials, but had a much greater effect for COUNT trials. The search time data for these two interactions can be seen in Figure 7.

4.0 DISCUSSION

The evoked pupillary response was sensitive to information processing demands in a visual search task. In particular, larger pupillary diameter was observed in the SEARCH condition where subjects were actively processing information relevant to task performance, as opposed to the CONTROL condition where subjects passively viewed the displays. However, the large baseline difference between the SEARCH and CONTROL conditions may only have indicated that subjects were more aroused in the active search task than in the psuedo-search task. In fact, many subjects complained of boredom and fatigue in the psuedo-search task.

Of greater import was that larger pupillary diameter, corresponding to longer search time, was observed for noncoded than for color-coded displays in the SEARCH condition. The Condition by Color Coding interaction for the D2-C difference component indicated that this effect was not an artifact of intensity differences between the color and monochrome displays or a result of the color displays having greater stimulatory value than the monochrome displays simply because they activated more photoreceptors. If pupil diameter was determined solely by some physical dimension of the displays, the same type of response would have been elicited in both the SEARCH and CONTROL conditions. Instead, pupil diameter was larger to the color displays

in the CONTROL condition, presumably because they were intrinsically more interesting than the monochrome displays.

The only effect of the display density manipulation was the Density by Color Coding interaction for the D2 component. This interaction was probably due to our procedure of terminating data collection at display offset along with the motor response. This procedure could have resulted in truncating the D2 component in the low-density color-coded condition when the trial was very easy and, consequently, response time was very short. Alternatively, D2 resolution may not have been completed in some high-density noncoded trials, particularly when the trial was very difficult and subjects did not complete their search within the 6-sec limit. Because of our procedure, it was unclear precisely how display density affects the pupillary response. It is clear, however, that task difficulty (at least as manipulated by color coding) interacts with display density to determine maximal pupil dilation.

In summary, these data indicate the potential usefulness of pupillary responses in evaluating the information processing requirements of visual displays. However, because our task was originally designed to evaluate visual search behavior, and not pupillary responses, several methodological deficiencies limited the conclusions that can be drawn from the data. We are currently in the process of adapting the visual search paradigm to the examination of pupillary responses in order to conduct further research in this area. The promise of the approach lies in the separation of the impact of some of the multiple determinants of mental workload.

5.0 REFERENCES

1. Wickens, C.D., & Derrick, W.: Workload measurement and multiple resources. Proceedings, 1981 IEEE conference on cybernetics and society. New York: Institute of Electrical and Electronics Engineers, 1981.

2. Carroll, J.B.: Individual difference relations in psychometric and experimental cognitive tasks. Report No. 163. Chapel Hill, NC: Thurstone Psychometric Laboratory, 1980.

3. Wickens, C.D., Sandry, D., & Vidulich, M.: Compatibility and resource competition between modalities of input, central processing, and output: Testing a model of complex task performance. Human Factors, 25, 1983, 227-248.

4. Walrath, L.C., Backs, R.W., & Hancock, G.A.: Eye movement during visual search: Effects of information density, color coding and search type. Psychophysiology, 23, 1986, 469. (Abstract).

5. Beatty, J.: Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychological Bulletin, 91, 1982, 276-292.

6. Goldwater, B.C.: Psychological significance of pupillary movements. Psychological Bulletin, 77, 1972, 340-355.

7. Libby, W.L., Lacey, B.C., & Lacey, J.I.: Pupillary and cardiac activity during visual attention. Psychophysiology, 10, 1973, 270-294.

8. SAS Institute Inc.: SAS Users Guide: Statistics, 1982 Edition. Cary, NC: SAS Institute Inc, 1982.

9. Winer, B.J.: Statistical Principals in Experimental Design. New York: McGraw Hill, 1962.

Figure 1. Pupil diameter data collection.

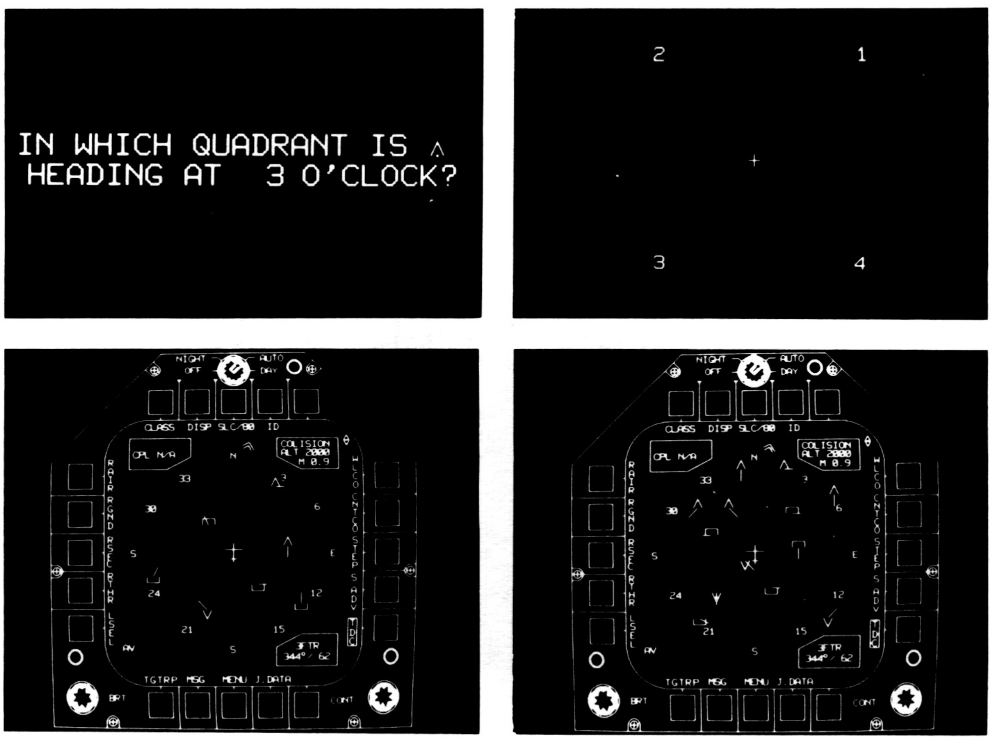


Figure 2. Examples of a question screen from the count condition (upper left), the calibration screen (upper right), a high-density display (lower right), and a low-density display (lower left).

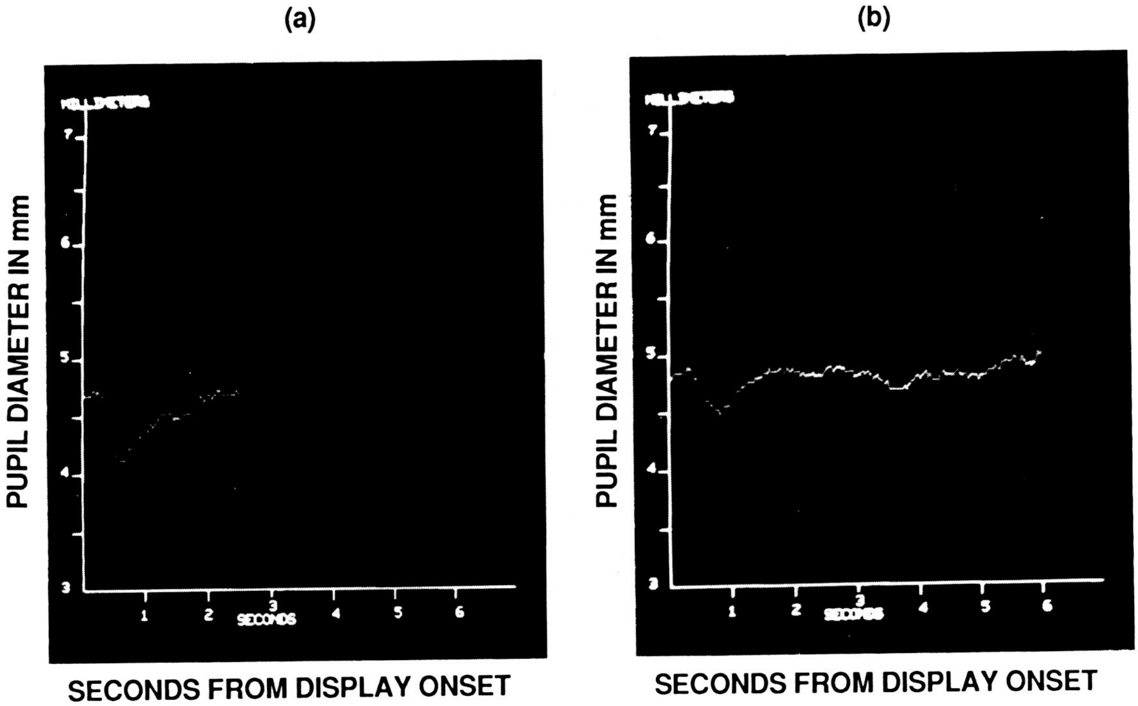


Figure 3. Illustrative single-trial pupillary responses from (a) color-coded, low-density, LOCATE and (b) noncoded, high-density, COUNT trials.

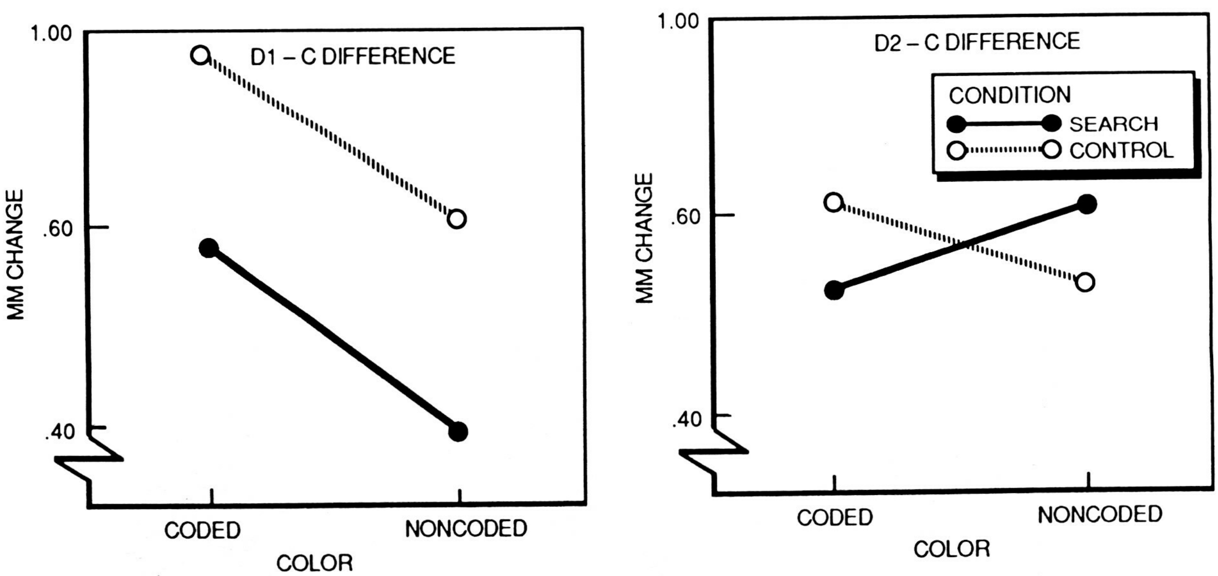


Figure 4. Color Coding and Condition effects for pupillary responses (n=8).

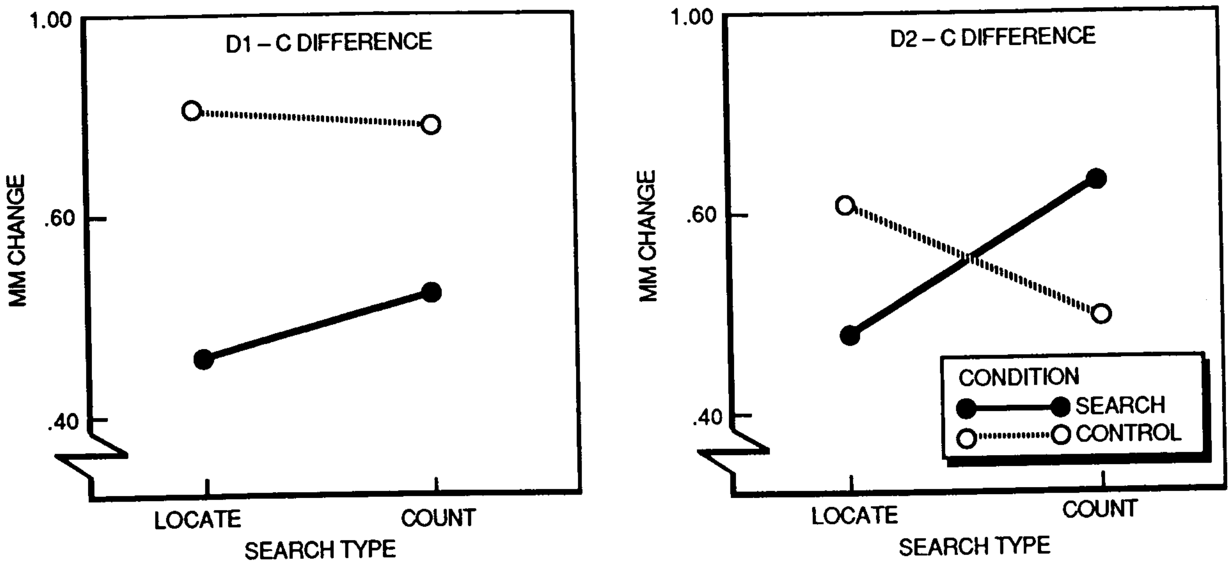


Figure 5. Search Type and Condition effects for pupillary responses (n=8).

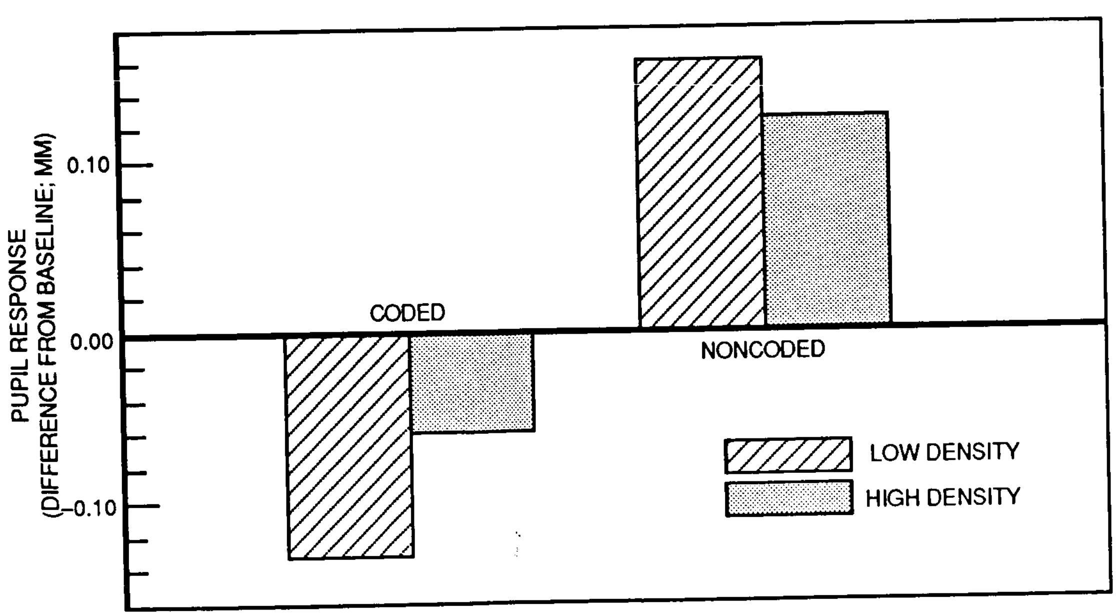


Figure 6. Density by Color Coding effect for the D2 pupillary response component (n=8).

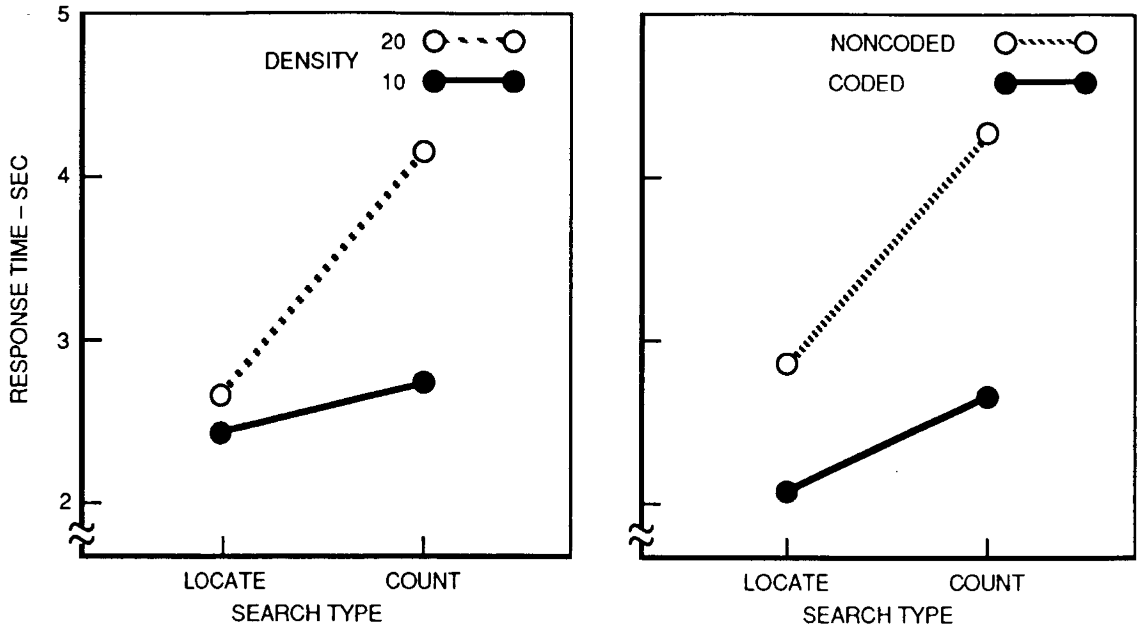


Figure 7. Effects of Color Coding, Density, and Search Type on response time (n=8).

N88-23379 S9-53
140797
8P

PROBE-EVOKED EVENT-RELATED POTENTIAL TECHNIQUES FOR EVALUATING ASPECTS OF ATTENTION AND INFORMATION PROCESSING

John A. Stern
Washington University Behavior Research Laboratory at
Malcolm Bliss Mental Health Center
St. Louis, Missouri

WG 032961

Probe event related potentials (Probe ERP) have been studied since the 1960s. In this context, a probe stimulus, by definition, is a stimulus irrelevant to task performance, which is introduced during task performance. The major premise underlying the use of probe ERPs is the belief that the ERP to such stimuli is significantly affected by task requirements. To the extent that the task and the probe stimulus share cerebral "space," one would hope that components of the probe ERP would be altered as a function of cerebral space being allocated to primary task performance.

Most of the recent literature dealing with probe ERPs has focused on the issue of hemispheric specialization. To the extent that one hemisphere is more utilized in the processing of a specific type of information, that hemisphere should demonstrate greater attenuation of probe ERPs than the less used hemisphere.

Earlier studies investigated more general questions, such as the effect of attention attracting visual stimuli on the ERP to light flashes. For example, Lehmann, Beeler, and Fender (1) and van Hof (2) report that when a patterned stimulus, as compared to a dark field or unpatterned stimulus, is presented to one eye, and flash stimuli to the other eye, the ERP to the flash stimuli is significantly affected. A structured target reduces the amplitude of the flash evoked response, as measured by the area under the ERP curve.

The major problem with many of the early studies was the lack of control for attentional variables. Could the described effects have simply been due to alterations in attention produced by the introduction of a task superimposed on the probe stimuli? The literature also is confusing, with respect to the nature of the response to probe stimuli. Some studies find augmentation; others, reduction; and still others, no effect as a function of primary task performance. Some studies report that early components of the ERPs are affected; others report late components are affected.

Since the mid-1970s, a number of laboratories have used probe ERP procedures to tap differential hemispheric processing. Galin and coworkers (3); Shucard and collaborators (4); and Papanicolaou and his collaborators (5) were some of the earliest investigators to utilize probe ERPs for the evaluation of differential hemispheric processing. The results of these investigations are generally supportive of hemispheric differences in information processing, as indexed by alterations in components of the probe ERP.

We will not review the results of these studies, but, in general, we concur with the critical comments made by Gevins and Schaffer (6) with respect to these and other studies purporting to demonstrate EEG correlates

of higher cortical functions. The following quote from their (1980) paper will alert the reader to the caustic nature of their comments.

"In the ensuing 50 years (since Berger's discovery of the human EEG, in general, and the alpha rhythm, in particular), no clear understanding of the relationship between EEG patterns and higher cortical functions has developed, despite an ever-increasing sophistication in experimental and analytic procedures" (p. 113). We do not propose to either critique their comments or accept them, carte blanche.

Which of their comments are most appropriate, with respect to the evaluation of studies utilizing probe ERPs? Most such studies have subjects performing relatively complex tasks, such as solving arithmetic problems, assembling Kohs blocks, or reading. Probe stimuli are presented at either fixed or random time points, while subjects are engaged in these tasks. Fixed, here, means that they are presented at regular time intervals, and become predictable on that basis. They are not, however, fixed with respect to either primary task stimulus presentation or task processing requirements. Random presentation, here, simply refers to temporal randomness, with respect to primary stimulus presentation. Many of these studies have concerned themselves with differences in these probe ERPs between bilaterally symmetrical skull sites, with the assumption that certain tasks principally tap the functions attributable to one hemisphere, while other tasks are more demanding of the other hemisphere. It is the contention of Gevins and Schaffer that there are no tasks which truly differentially tap the two hemispheres, and that the performance of any task involves dynamic processes that are not restricted to one or the other hemisphere. Gevins et al. (7, 8) present data that even simple perceptual tasks involving spatial judgment and visuomotor integration, produce complex patterns of cortical activity with shifts not only between hemispheres, but also within a hemisphere. These are, truly, variable spatio-temporal events.

During this complex interplay between various cortical and subcortical sites, we now introduce probe stimuli. These come at essentially random time points during such information processing. How can they possibly provide us with much coherent information? The answer is that, for every published study which has obtained positive results, there is at least one published study with negative results, as well as untold studies with negative or inconclusive results.

We will briefly review the results of one study purporting to demonstrate laterality effects on probe ERPs attributable to differential processing of an arithmetic and a visuospatial processing task. We have not singled out this study, but have selected it randomly from those available to us (Papanikolaou, 5).

Probe stimuli, in this experiment, were 70 dB., 1000 Hz tones, presented at a rate of 1.3 per second, with 84% of the tones 50 msec in duration, and 16% of 60 msec duration.

The primary experimental task involved the visual presentation of a random shape, and a shape divided into three irregular sections. A number (between one and nine) was centered in each irregular section, as well as in the full random shape. For 84% of the trials, the three segments, when

joined, matched the random shape and the sum of the three numbers in the sections, when added together, matched the number in the full shape.

ERPs were recorded from the temporal and parietal areas on the left and right side (T_3, P_3, T_4, P_4), using linked ear reference. Under the control condition, subjects were required to gaze at the visual display, but attend to the tones and make a simple manual response to the 60 msec tone. In the two experimental conditions, the subject was required to either attend to the shapes or the numbers, and make the same simple, manual response to the "aberrant" stimulus set.

N1 (90 msec latency) - P2 (170 msec latency) amplitude difference was the component of interest. This measure was obtained for all three conditions, and a ratio of N1-P2 amplitude, with the control task as the denominator and the experimental tasks as the numerator ("arithmetic" and "visuospatial"), were calculated. The results were that this ratio was less than 1.0 for the visuospatial task, regardless of recording site, while there was some augmentation for the arithmetic task for P_3, P_4, and T_4. For P_3 and P_4, the augmentation was approximately 7%, while for T_4, it was 15%. Using "t" tests to evaluate whether these ratios were significantly different from Ø, none of the augmenting proved to be reliable. Significant attenuation was obtained at T_3 (10%, 15%) for both tasks, while significant attenuation was also found at P_3 (10%) and P_4 (25%) for the visuospatial task only. Attenuation for the visuospatial task was significantly greater at P_4, as compared to P_3 and T_3. These results, (ref. 5, p. 287, last paragraph) were interpreted as follows:

> These findings reaffirm the widely documented involvement of the left temporal area of dextral individuals in serial-analytic operations such as those required by the present arithmetic task. They also accord with the notion of predominant contribution of the right posterior region of the brain in visuospatial processing (e.g., see Hecaen & Albert, ref. 9). In addition, however, they indicate that the left, rather than the right, temporal area was involved in that task. At present, it is unclear whether this pattern of cerebral excitation, especially the involvement of the left temporal area, is representative of visuospatial processing at large, or confined to the specific task employed in this study. In this task, two alternative strategies could be used equally efficiently: The first would require mental segregation of the scattered sections and subsequent comparison of the resulting shape to the intact one. The second could simply involve comparison of each scattered section to the sections of the intact shape. Though both strategies require visuospatial processing, the latter does not entail mental manipulation of the visual stimuli and it does contain a serial-analytic component. Whether employment of this strategy accounts for the observed engagement of the left hemisphere in the present study, is a question deserving further exploration.

Our critique of this study focuses on two major issues, one dealing with: a) the logic of the specific control condition used, and b) the logic of introducing probes at 1.3 sec intervals during information acquisition processing and responding.

a. Why would one use a condition in which subjects are required to process information presented in the auditory mode as a control

for evaluating the ERP to that same stimulus condition, where the S is (attending to visually displayed information), not attending to the auditory information. Thus, perhaps a more reasonable control might have been a condition where everything was presented without any task demands.

b. The logic of averaging across probe stimuli presented at four different points in time, with respect to primary task performance, is also suspect.

Since visual stimuli were presented for four seconds, and the auditory stimuli at 1.3 second intervals, one may infer that auditory stimuli were presented concurrent with visual stimulus onset, 1.3 and 2.6 seconds into the visual stimulus presentation period, and immediately preceding termination of that stimulus. Evoked responses to these stimuli were averaged.

If one conceives of the auditory discrimination, the arithmetic and the visuospatial tasks as information processing tasks, then what happens during the four second stimulus presentation period must differ from second to second, or millisecond to millisecond. For the auditory task, the presentation of the visual stimulus signals the onset of a series of four tones. Most of these tones (84%) are 50 msec in duration. The subject must discriminate between 50 and 60 msec duration stimuli and make a manual response to the 60 msec tone pip. This involves the development of an internal "model," for the shorter of the two stimuli, and deciding that the longer one does not match that model. (We suspect that the model should be for the shorter stimulus, because it is more frequently presented). Under these conditions, we would not expect any eye movements. This expectation has some empirical foundation, albeit utilizing stimuli of longer durations. For the arithmetic task, he may sequentially scan the three partial displays, abstract the numbers, add them together, and then look at the full display and compare that number with his addition, and make the appropriate response. For the visuospatial task, he probably scans back and forth between the segments and the full figure to make the decision. Thus, there is considerably more visual scanning activity in the latter task than in the arithmetic task, and more visual scanning in the arithmetic than the auditory discrimination task. One might also suspect that the time necessary to arrive at a decision might differ between the two (or even three) tasks, and that the timing of the motor response may affect the ERP.

We are, thus, surprised that significant results were obtained in this study. I am not surprised that the results were interpretable. One of man's unique abilities is the generation of hypotheses to account for any set of results. I can rationalize almost any set of data involving CNS activity, if you will allow me the concepts of excitation and inhibition.

In view of these rather negative and devastating comments, what is it that we did which we consider a marked improvement over the approaches taken by other researchers utilizing probe ERPs? It is our contention that probe ERPs have to be presented at points in time where one can be assured that more or less specific aspects of information intake or processing are occurring. Thus, we time-locked our probe ERPs to aspects of stimulus presentation. Such time-locking has been relatively crude, and can be improved upon in a number of ways.

Before suggesting such improvements, I will review results of a study conducted in our laboratory utilizing such probe ERPs, as well as evaluating ERPs to primary task performance. These studies have also evaluated other physiological measures, specifically, heat rate (HR) and aspects of blinking (10).

We have modified the Sternberg memory paradigm to allow us to evaluate aspects of anticipation or expectancy, information acquisition and retention, or memory and comparison. First of all, our procedure provides the subject with information about the expected memory set (e.g., is it small or large; does it involve symbol set A or symbol set B). Second, since a fixed time is allowed to elapse between presentation of this CUE information and presentation of the MEMORY set, he also "knows" when the memory set presentation will occur. The CUE stimulus, thus, provides him with up to three units of information about the upcoming memory set, size, nature, and its time of arrival. The MEMORY set is then presented for a fixed time period, followed by a constant duration retention period. Following this, a TEST stimulus is presented, which is or is not a member of the set presented during the MEMORY period. The subject makes a discriminative response. After a fixed interval, the next CUE stimulus is presented. All information is visually presented and is under computer control. In addition to these information bearing stimuli, the subject is presented a probe stimulus, which occurs at one of six temporal locations--three between CUE and MEMORY sets, and three between the MEMORY and TEST stimuli. Probe stimuli occurred early in the middle, or immediately preceding presentation of the next stimulus. In the first experiment, early was defined as 1300 msec following stimulus offset, middle was 2500 msec after offset,and late was one second before presentation of the next stimulus.

We evaluated the ERPs to these probe stimuli, as well as the CUE, MEMORY, and TEST stimulus. With respect to the latter stimuli, what did we learn?

1. Knowing what to expect, whether it involved partial or full knowledge, leads to smaller P3 amplitude to the MEMORY stimuli, than not knowing what to expect. This effect is restricted to the anticipation of large set size only (Bauer, 1987, ref. 10) (Donchin, 1981, ref. 11) expected stimuli elicit smaller P3 than unexpected ones).

2. CUE and MEMORY stimulus produced ERP differences for P1, P2 and P3. P1 and P2 amplitudes are larger to the CUE stimulus; P3--amplitude is greater for MEMORY set.

3. With respect to the memory set, we find:

   a. P3 amplitude directly related to set size, with the larger set size generating larger P3's than the smaller set sizes (two studies).

   b. We found the P2 amplitude component of the ERP to the memory set significantly greater on the left side of the head (P_3) for English, as compared to Katakana characters. It was significantly greater on the right side of the heads (P_4) for Katakana characters.

c. This effect occurs only under fully cued conditions. In other words, the laterality effect to the MEMORY set only occurs when subjects fully know what to expect (and can "prepare" to deal with the material).

4. To the TEST stimulus, we

a. corroborated both previous results from our and other laboratories, in that P3 amplitude is inversely related to set size. This effect is seen equally in left and right derivations.

b. found N2 amplitude to increase with set size on match trials only, and this only over the right hemisphere.

What results have we obtained from our probe ERPs, to date?

In our first study (Pz, Fz), we demonstrated that differential effects of set size were restricted to the probe which immediately preceded presentation of the MEMORY set and the probe immediately following the MEMORY set. Amplitude of the P1-N1 component increased with set size in anticipation of the MEMORY set and N1-P2 decreased with increasing set size immediately following memory set size presentation.

The "anticipatory" effect appears to be limited to midline lead placements, since it was not replicated in a study in which we recorded from parietal and temporal leads on the left and right sides.

For the MEMORY period (P), there was a significant probe position effect in the Bauer study, with both P1-N1 and N1-P2 increasing in amplitude, as one moved from the first to the third probe position, and a decrease in P2-N2.

Although I continue to have lingering doubts about the applicability of ERPs in simulation and real world environments, our studies, to date, have provided us with some landmarks suggesting both the utility of primary and probe stimuli on both probe and primary task elicited ERPs in the evaluation of "spare channel capacity."

My lingering doubts are not restricted to the application of the ERP to simulation and field condition, but to the laboratory situation, as well. Relatively minor changes in the experimental paradigm can produce major shifts in ERP findings. Whether this is interpreted as sensitivity of the ERP paradigm, or whether one attributes the ERP results to error variance, is a highly subjective matter.

A recently published study by Brumaghim and collaborators (1987, ref. 12) demonstrates such changes in ERP components nicely. They restricted their analyses to the P_{3b} component as affected by methyphenidate, and conducted two studies. In both studies, they found P3b latency affected by memory load, but in only one of the two studies was it affected by methylphenidate. To quote "The explanation of this effect, however, is not clear (p. 371). I suspect that everyone doing ERP research can come up with some examples of non-replicability of results from one study to another).

In spite of my doubts, how might one go about the task of using task elicited ERPs in the flight simulator. If, for example, we can take time of arrival of the eyes on a particular instrument as one variable of concern, and dwell time on the instrument as a second variable, one which reflects importance of the information displayed, one might look at ERPs triggered by saccade termination (the one which slews the eyes to the appropriate instrument) for fixation pauses of specified durations. One might go a step (or two) further, and look at patterns of ocular activity and associated ERPs.

If looking at instrument A is followed by looking at instrument B, assign the ERP to a different bin than if the second look is on instrument C, D, or E. It may well be that the importance of the information obtained from display A is greater, if followed by a glance at B, than any other location, and that the ERP to momentarily "important" display will be different from that elicited by a routine instrument check. With respect to probe ERPs, one could consider the introduction of such probes associated with the eyes falling on a particular display. Is the probe ERP to a display from which information is abstracted rapidly discriminable from one where such information abstraction is slow?

Thus, both primary stimulus, as well as probe ERPs, can be moved from the laboratory to the simulator, and to field conditions.

REFERENCES

1. Lehmann, D., Beeler, G.W. Jr., and Fender, D.H. EEG responses to light flashes during the observation of stabilized and normal retinal images. EEG and Clinical Neurophysiology, 1967, 22, 136-142.

2. van Hof, M.W. The relation between the cortical responses to flash and to flicker in man. Acto physiologica pharmacologia neederlander, 1960, 9, 210-240.

3. Galin, D. and Ellis, R.R. Asymmetry in evoked potential as an index of lateralization of cognitive processes: Relation to EEG alpha asymmetry. Neuropsychologia, 1975, 13, 45-50.

4. Shucard, S.W., Chucard, J.S. and Thomas, D.F. Auditory evoked potentials as probes of hemispheric differences in cognitive processing. Science, 1977, 212, 344-346.

5. Papanicolaou, A.C., Schmidt, A.L., Moor, B.D. and Eisenberg, H.M. Cerebral activation patterns in an arithmetic and visuospatial study processing task. International Journal of Neuroscience, 1983, 20, 283-288.

6. Gevins, A.S. and Schaffer, R.E. A critical review of EEG correlates of higher cortical functions. CRC Critical Review in Bioengineering, 1980, 113-164.

7. Gevins, A.S., Doyle, J.C., Cutillo, B.A. et al. Electrical potentials in human brain during cognition: New method reveals dynamic patterns of correlation. Science, 1981, 213, 918-922.

8. Gevins, A.S. Brain potential evidence for lateralization of higher cognitive functions. In J.P. Hellige (ed) Cerebral Hemisphere Asymmetry Praeger, 1983.

9. Hecan, H. and Albert, L.M. Human Neuropsychology, New York, Wiley, 1978.

10. Bauer, L.O., Goldstein, R. and Stern, J.A. Effects of information processing demands on physiological response patterns. Human Factors, 1987, 29.

11. Donchin, E. Surprise!... Surprise? Psychophysiology, 1981, 8, 493-513.

12. Brumaghim, J.T., Klorman, R., Strauss, J., Lewine, J.D. and Goldstein, M.G. Does methylphenidate affect information processing? Findings from two studies on performance and P3b latency. Psychophysiology, 1987, 24, 361-372.

N88-23380

5/0-53
40798
24P.

STEADY-STATE EVOKED POTENTIALS POSSIBILITIES FOR MENTAL-STATE ESTIMATION

Andrew M. Junker
Armstrong Aerospace Medical Research Laboratory
Wright-Patterson AFB, Ohio

John H. Schnurer, David F. Ingle, and Craig W. Downey
Systems Research Laboratories
Dayton, Ohio

AG627259

S9628092

ABSTRACT

The use of the human steady-state evoked potential (SSEP) as a possible measure of mental-state estimation is explored. A method for evoking a visual response to a sum-of-ten sine waves is presented. This approach provides simultaneous multiple frequency measurements of the human EEG to the evoking stimulus in terms of describing functions (gain and phase) and remnant spectra. Ways in which these quantities vary with the addition of performance tasks (manual tracking, grammatical reasoning, and decision making) are presented. Models of the describing function measures can be formulated using systems engineering technology. Relationships between model parameters and performance scores during manual tracking are discussed. Problems of unresponsiveness and lack of repeatability of subject responses are addressed in terms of a need for loop closure of the SSEP. A technique to achieve loop closure using a lock-in amplifier approach is presented. Results of a study designed to test the effectiveness of using feedback to consciously connect humans to their evoked response are presented. Findings indicate that conscious control of EEG is possible. Implications of these results in terms of secondary tasks for mental-state estimation and brain actuated control are addressed.

INTRODUCTION

By using appropriate signal averaging techniques, it is possible to detect a response in the human electroencephalograph (EEG) to evoking stimuli. When the stimulus is sinusoidally modulated the result is called a steady state evoked potential (SSEP). Research in this area (Spekreijse, 1966; Regan, 1972; Wilson and O'Donnell, 1980) suggests that the SSEP may be a useful indicator for mental-state estimation.

Using a light stimulus modulated by a sum of sine waves,

a steady state evoked potential can be elicited that contains responses at all of the component frequencies of the driving stimulus. A technique has been developed to drive the stimulus with a 10 frequency sum of sines. This technique has been refined and the analysis has been upgraded to a level of sophistication that allows detailed analysis to be applied to the discrete Fourier transforms of the SSEP and the evoking stimulus. This analysis simultaneously produces describing function measures and background EEG spectra (Junker et. al., 1987). The describing function provides gain and phase information as a function of stimulus frequency, measures which are systems engineering based. The background EEG spectrum, referred to as the remnant in this report, provides information about the average power adjacent to, but not including the power at, stimulus frequencies. Thus, this remnant represents an average measure of EEG activity excluding the linear response to the evoking stimulus.

This analysis has been applied to SSEPs in taskloading and non-taskloading conditions. The tasks used were manual tracking, grammatical reasoning and decision making.

METHODOLOGY

The experimental apparatus used to obtain SSEP measures is illustrated in Figure 1. The apparatus consists of a stimulus presentation device which simultaneously delivered the evoking stimulus (flickering light) and a video task display. This presentation was achieved by combining the two images via a half-silvered mirror at 45 degrees to each image. The evoking stimulus was produced by two fluorescent light tubes behind a diffusing screen which distributed the light over the entire visual field. The intensity of the light was measured by a photocell placed at the subject's viewing point. The tasks were displayed on the video monitor. The average intensity of the evoking light was sufficiently low that a subject could comfortably discern the video task display within the same visual field.

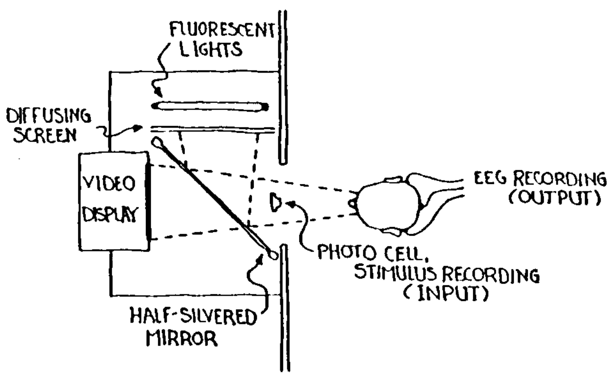


Figure 1. Experimental apparatus.

Subjects were seated in a darkened chamber facing the test apparatus. For the task conditions subjects were instructed to concentrate on the tasks. At the end of each 90 second trial, the subject's performance score appeared on the screen. For the non-task condition, called lights only, subjects were instructed to "relax and fixate on the center of the screen". Sessions were limited to 20 trials.

The EEG was recorded with silver/silver cloride electrodes at Oz with the right mastoid as reference and left mastoid as ground for the manual tracking. The grammatical reasoning and modeling results are reported here. For the investigation of decision making effects and loop-closure, gold cup electrodes were used with O1 as signal, P3 as reference and right ear as ground. Sum-of-sines generation and data collection were accomplished on a PDP 11/60 computer. The two channels of data (photocell and EEG) were filtered, digitized and stored for analysis. The collected data were discrete Fourier transformed, ensemble averaged, describing functions and remnant were computed, and the results were then plotted. Estimates of mean values for the gain and phase computations across trials were computed. For an indication of mean variability, standard errors were computed. The describing function gain (amplitude ratios of the EEG to photocell) indicates evoked response sensitivity at the component frequencies. The phase values relate to neurophysiological dynamics and transmission latency between photocell and EEG measurement.

Three tasks, requiring various levels of visual, mental, and motor processing, were used to elicit diverse cognitive states with the intention of evoking different visual-cortical responses. The three tasks were similar in that the input came from the video display and the output from subjects was produced by manual operation of a control stick or push-buttons.

The manual tracking task involved control of a first order instability driven by pseudo-random noise. Visually this involved minimizing a displayed error by keeping a cursor superimposed upon a moving dot. This task required continuous manual control and little or no conscious decision making once the task had been learned (Zacharias and Levison, 1979).

A grammatical reasoning task was used which imposed variable processing demands on mental resources used for the manipulation of grammatical information (Shingledecker et. al., 1983). Stimulus items were two sentences of varying syntactic structure accompanied by a set of three symbols. The sentences had to be analyzed to determine whether they correctly described the ordering of the characters in the symbol set.

The decision making task involved the problem of allocating attention among multiple tasks in a supervisory control system (Pattipati et. al., 1979). Subjects observed the video display on which multiple concomitant tasks were represented by moving rectangular bars. The bars appeared at the left edge of the screen and moved at different velocities to the right, disappearing upon reaching the right edge. At any given time there were, at most, five tasks displayed with a maximum of one on each line. The subjects could process a task by depressing the appropriate push-button. Once a button had been pushed, the computer remained dedicated to that task until task completion or the task ran off the screen. By processing a task successfully, the subject was credited with the corresponding reward, and the completed task was eliminated from the display. Two levels of difficulty were used. In the "easy" condition it was possible to successfully allocate attention among the multiple tasks. In the "hard" condition the time required exceeded the time available and it was not possible to complete all allocations successfully.

The sum-of-sines stimulus was composed of 10 harmonically non-related multiples of the fundamental frequency of 0.0244 Hz. In addition, none of these component frequencies contained a sum or difference of any of the other component frequencies. This restriction on the sine wave frequency selection was implemented to avoid first order nonlinear interactions. The component frequencies ranged from approximately 6.25 to 21.74 Hz, with intermediate frequencies at 7.73, 9.49, 11.49, 13.25, 14.74, 16.49, 18.25, and 20.23 Hz. For every data collecting trial, starting phase values for each of the 10 component sine waves were randomized, ensuring that the time sequence of flickering light presentation was random from trial to trial. By utilizing randomized starting phase values with the summing of the 10 sinusoids a peak depth of modulation of 13 % per sinusoid was possible. Results for two levels of depth of modulation (6.5% and 13%) and two levels of average luminance, (40 foot-Lamberts, (ftL), and 80 ftL) are presented. For a detailed discussion of the rationale for designing sum-of-sines inputs the reader is referred to Junker et. al., 1987.

STIMULUS EFFECTS

Investigation into the effects of stimulus parameter characteristics is perhaps best summarized in Figures 2 and 3. For the subjects tested, the evoked response frequencies of greatest sensitivity were between 9.49 Hz and 18.25 Hz. Two areas of obvious sensitivity were the alpha band and beta band. For the lowest level of modulation and intensity, and thus stimulus power, a strong response was evoked at 9.49 Hz and a not so strong (but obvious) response occurred at 16.49

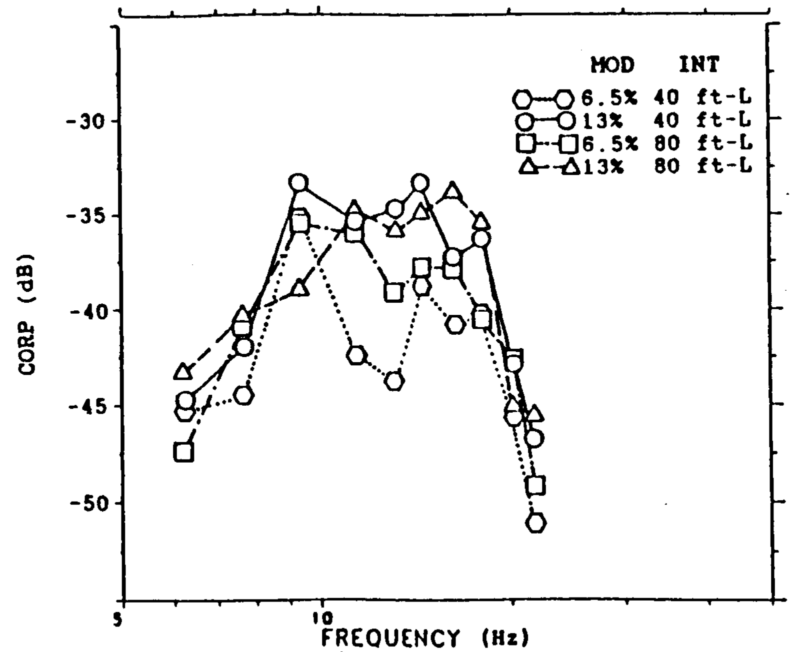


Figure 2. Effects of stimulus parameters; MODulation, and INTensity, on SSEP correlated power (power at evoking stimulus frequencies).

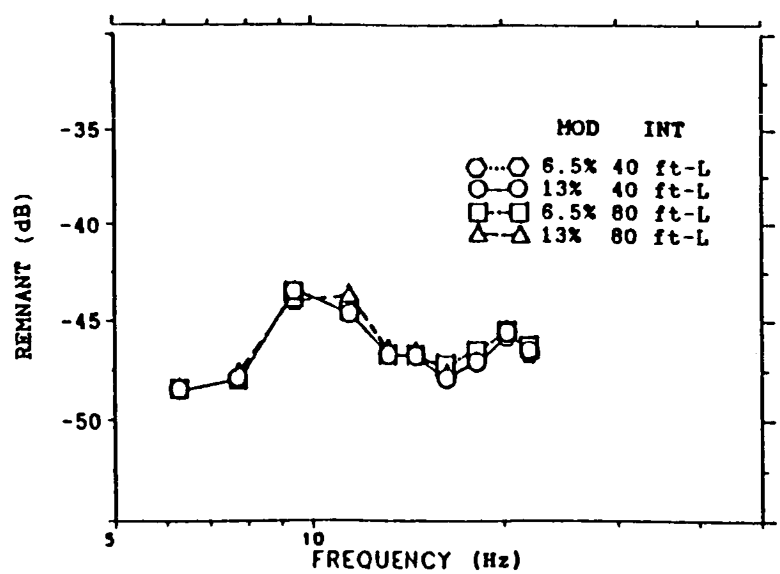


Figure 3. Effects of stimulus parameters; MODulation, and INTensity, on background EEG or remnant.

Hz. Increasing the depth of modulation to 13%, with the intensity unchanged (40 ftL), resulted in the largest evoked response and a flattening in the correlated EEG power spectrum (11.49 Hz to 14.74 Hz). At 13% modulation, increasing the intensity further (to 80 ftL) succeeded only in producing a slightly noticeable increase in the evoked response at 16.49 Hz. This high level of intensity and modulation actually resulted in the smallest evoked response at 9.49 Hz. These results indicate that the evoked response is a function of frequency as well as stimulus strength. These findings correspond to others reported in the literature (Regan 1972). It was also observed that saturation across frequencies was unequal, the alpha region being the most sensitive.

As can be seen in Figure 3, the remnant responses were only mildly affected by the different stimulus parameter values. In addition it can be observed that the alpha peaking in the remnant curves corresponded to the alpha sensitivity in the evoked responses of Figure 2. The results also indicated that differences in evoked responses between subjects were significant, and that they must be considered for a more complete picture of visual-cortical functioning.

From our results, it can be concluded that the lower level of intensity and higher level of modulation provide the better stimulus parameter values. In designing a stimulus, it would be best to choose values which cause minimal distraction of the tasks being investigated. An intensity level of 40 ftL was adequate for the experimental paradigm investigated for this report.

The investigation of stimulus parameters points to future research possibilities. Tailoring the stimulus spectrum to each individual as a function of their evoked response sensitivity may produce optimal SSEP responses.

TASK EFFECTS

Different effects upon the visual-cortical response were observed for the three tasks investigated. Manual tracking had the least effect for most subjects, and grammatical reasoning and decision making had the greatest effect.

Comparisons between lights only (LO), manual tracking (MT), and grammatical reasoning (GR) for 4 of the subjects tested are given in Figure 4. Results indicate that the more mental processing required, the greater the alpha band decrease and the greater the beta band increase. Of course this is somewhat specific to each subject tested. Subjects 02 and 05 could be classified as alpha responders due to their large alpha band remnant peaks (Figure 4a). For these subjects, with task loading, a decreasing remnant alpha

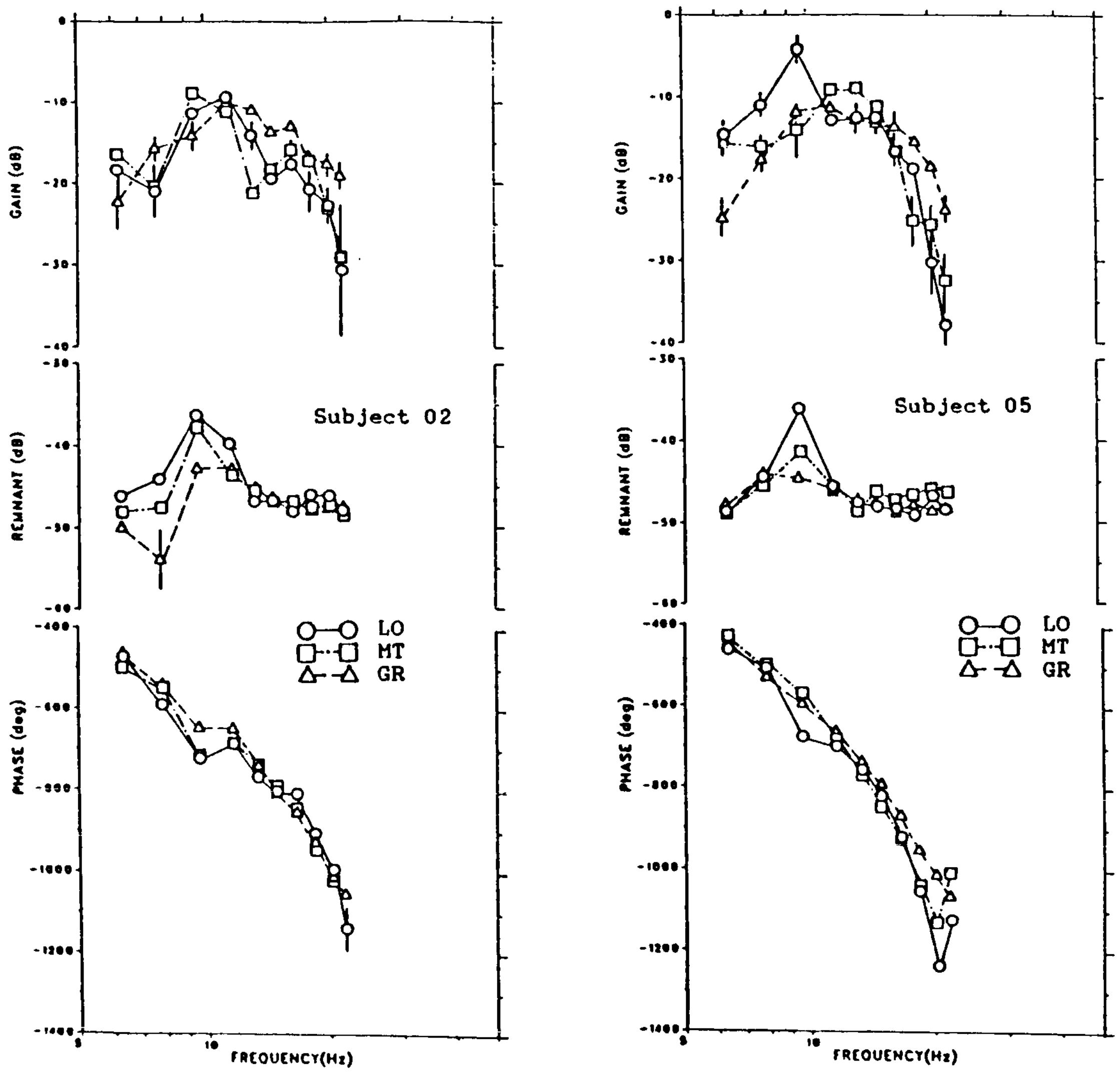


Figure 4a. SSEP describing functions (gain and phase) and remnant across three conditions; Lights Only (LO), Manual Tracking (MT), and Grammatical Reasoning (GR).

response corresponding to the degree of mental processing required can be seen. Subjects 10 and 15, non-alpha responders, do not exhibit such responses (Figure 4b).

Results from the decision making tasks on the SSEP are presented in Figure 5. During decision making as compared to the lights only condition, a consistent reduction in phase lag in the beta band was observed for all subjects tested (refer to Figure 5). As in the tracking and grammatical reasoning conditions, reductions in the alpha band and increases in the beta band with task loading could be observed. There were, however, no observable differences in

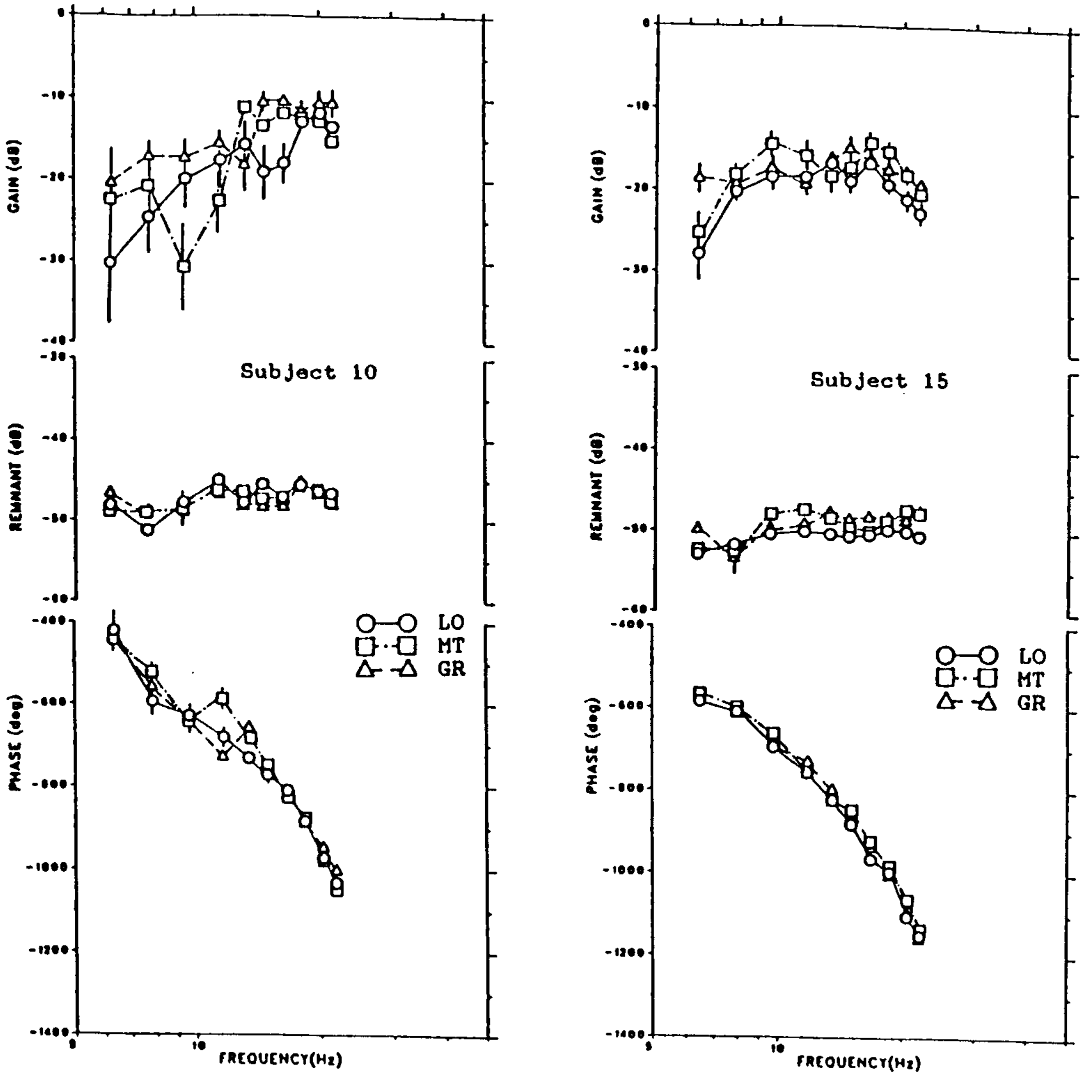


Figure 4b. SSEP describing functions (gain and phase) and remnant across three conditions.

the evoked responses across the two levels of decision making task difficulty. Subjects 13 and 77 could be classified as alpha responders based upon their remnant and gain responses in the alpha region (Figure 5a).

The changes across tasks were specific to each individual tested. The differences in subject responses suggest that it would be useful to group subjects into at least two groups: alpha responders, and non-alpha responders. Determination of how to group each subject could be based upon alpha band resonance or peak responses for remnant and gain. With task loading, subjects with alpha decreases in both the remnant and gain response could be classified as alpha responders. Non-alpha responders could be

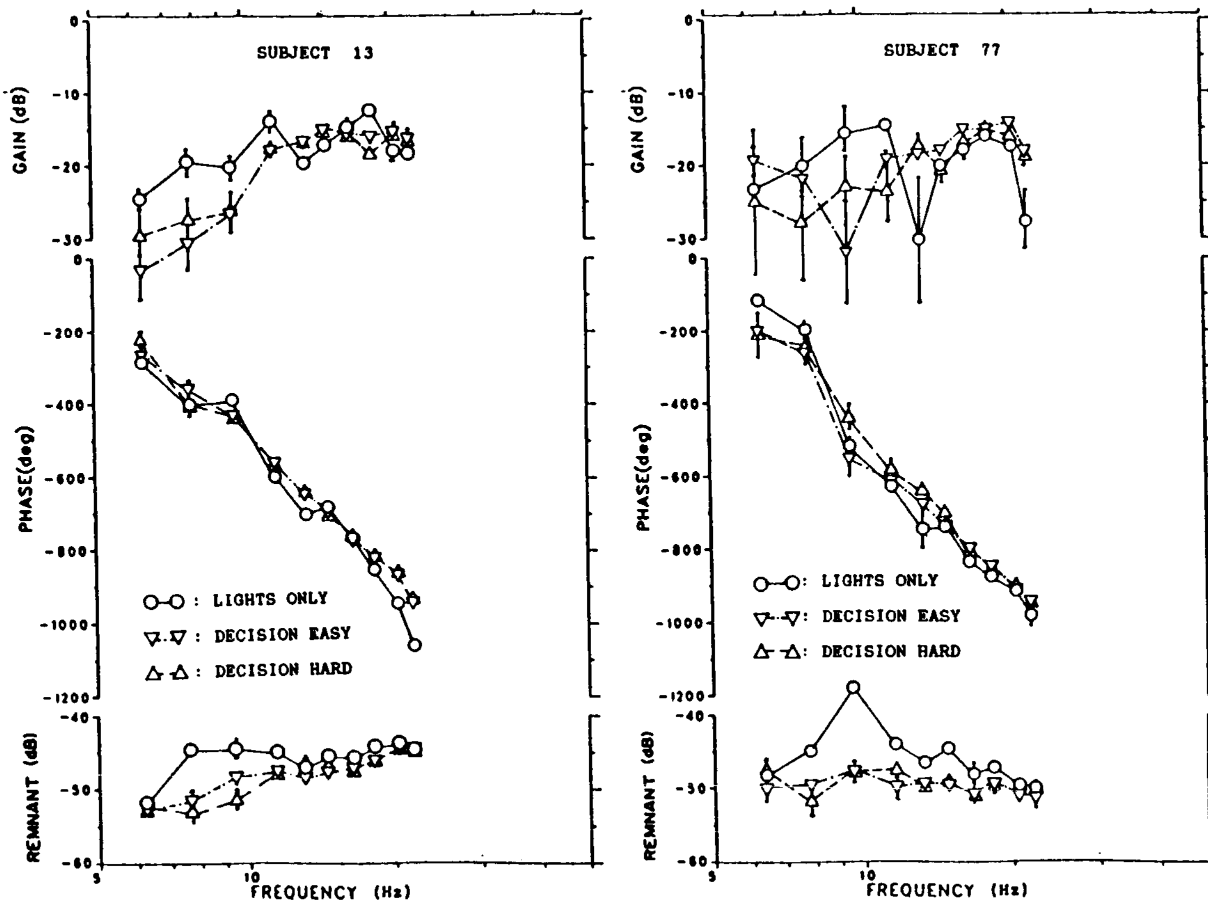


Figure 5a. SSEP describing functions and remnant for two levels of decision making task difficulty. Note large alpha response in lights only condition.

characterized primarily by a beta increase in gain and remnant with task loading.

Gain curve changes corresponded to remnant changes in the alpha band for subjects classified as alpha responders (Subjects 02, 05, 13, and 77). In the beta band (above 13 Hz) the gain curve activity appeared to be independent of the measured remnant for most subjects tested.

MODELING

Describing function data were modeled using a second order linear model form. Results of the model match (Figure 6) indicated that a good match could be achieved for some

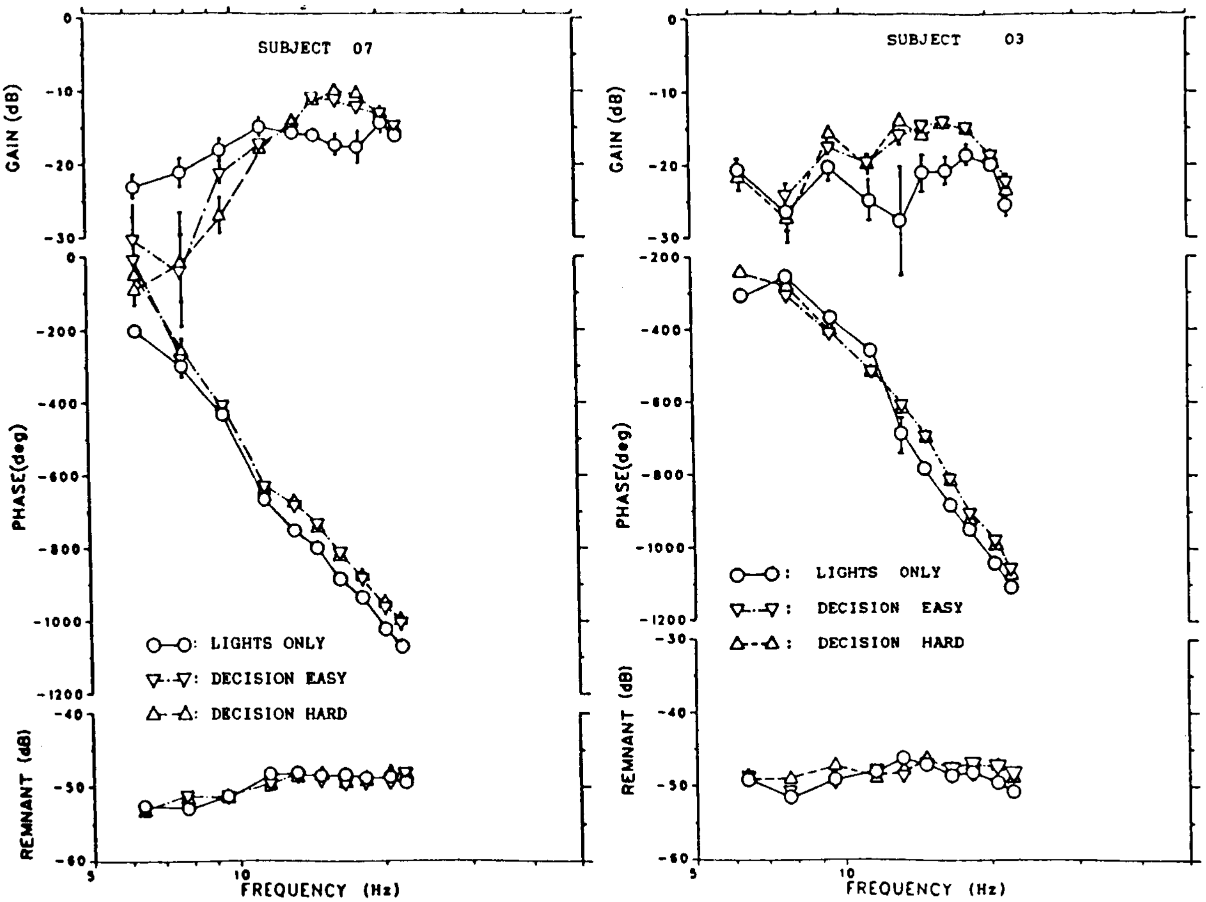


Figure 5b. SSEP describing functions and remnant, two levels of decision making. Note the absence of alpha changes in remnant from lights only to decision making.

subjects and not others. Due to individual differences in the evoked responses, it will be necessary to tailor the form of the model used to each subject. Perhaps by grouping subjects into two groups (alpha and non-alpha responders), two general model forms would be sufficient to compress the visual-cortical response data into a more parsimonious format.

A simple gain-delay model was useful as an aid in phase unwrapping. It was also used to parameterize the SSEP describing functions in terms of gain and delay. These values were compared to performance scores for the manual tracking task (refer to Table 1). Subject 10 achieved the best performance as indicated by the lowest error score, and Subject 15 achieved the worst as indicated by the largest

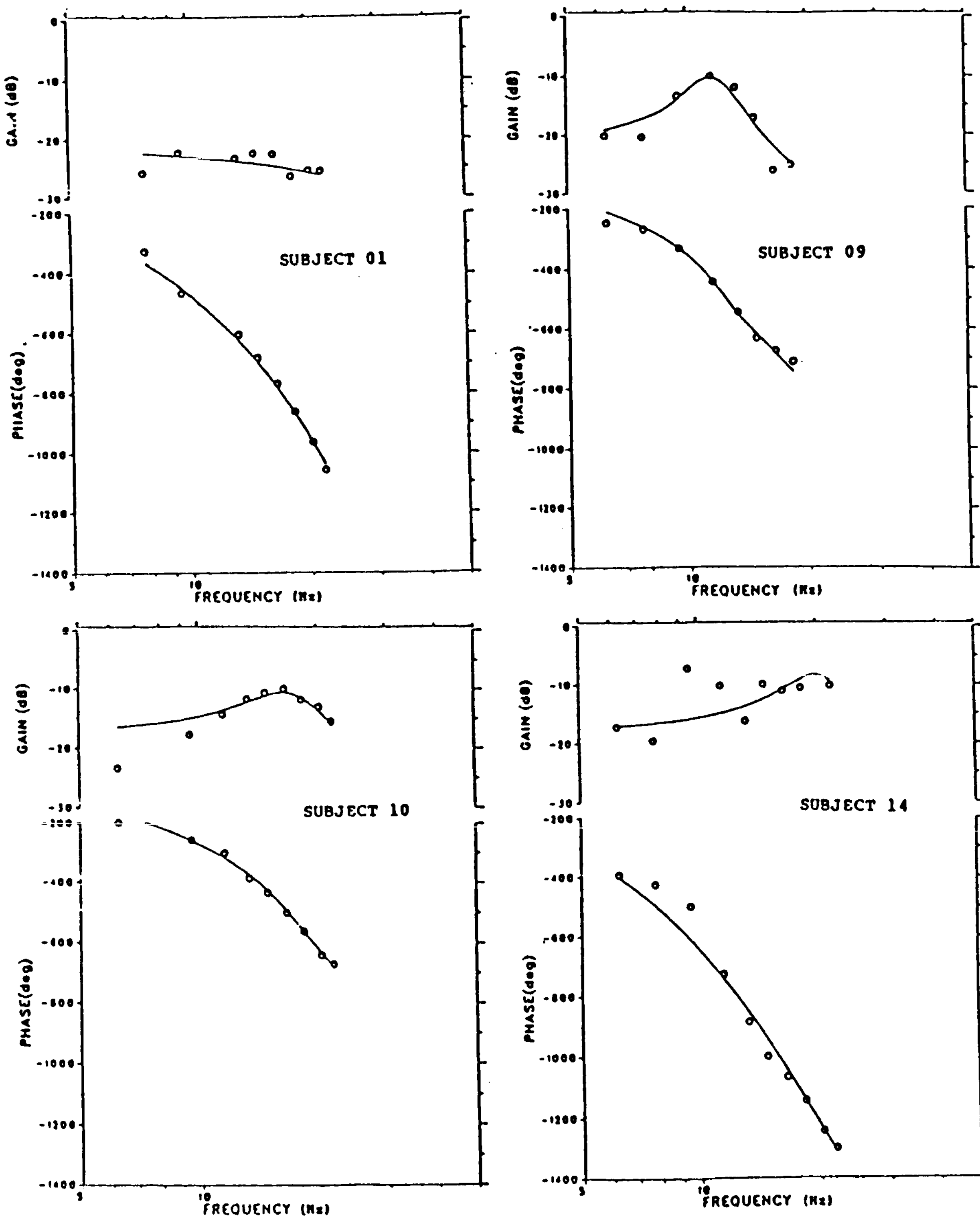


Figure 6. SSEP response, lights only condition. Circles represent phase-adjusted experimental values, lines are model predictions.

TABLE 1 Manual tracking performance scores and SSEP describing function model results (gain and delay values).

SUBJ #	RMS ERROR MEAN	SD	MODEL GAIN	DELAY
02	1.78	0.40	.151	.169
05	2.20	0.52	.240	.124
10	1.32	0.20	.222	.109
15	2.43	0.51	.135	.126

error score. It is interesting to note that Subject 10 also had the lowest modeled SSEP delay and Subject 15 had the lowest modeled SSEP gain. These results suggest the possibility that task performance may correlate with visual-cortical response frequency measures. Thus model parameterization may provide predictive information regarding a subject's ability to perform a particular task.

LOOP-CLOSURE OF THE VISUAL-CORTICAL RESPONSE

The results of our research effort indicate that describing functions can be obtained and that they are sensitive to changes in task loading. It was also found that the results are unique to each individual within the general classifications of alpha and non-alpha responders. Further, it was found that the results are sensitive to attention, especially in the alpha band.

These results are promising, however there is one difficulty with this and perhaps other evoked physiological measures that needs to be addressed. The visual-cortical response is an open loop measure. Unlike manual control, where an optimal behavior for best performance exists, the subject is not provided with an environment directing a certain response.

In the lights-only condition, subjects were told to "look at the lights". No feedback relative to how well they were responding was provided. Even with this lack of feedback or loop closure, the evoked response was somewhat repeatable. This is demonstrated in Figure 7 for two subjects that were tested over a 3 year span. It is interesting to note that task loading often increased the evoked response and reduced response variability. However, subjects were often unaware of their state of attention, resulting in a weak or unevoked response.

Based upon what was learned from manual control experiments (Levison, 1983; Levison and Junker, 1978; Levison et. al., 1971), it was concluded that the solution to

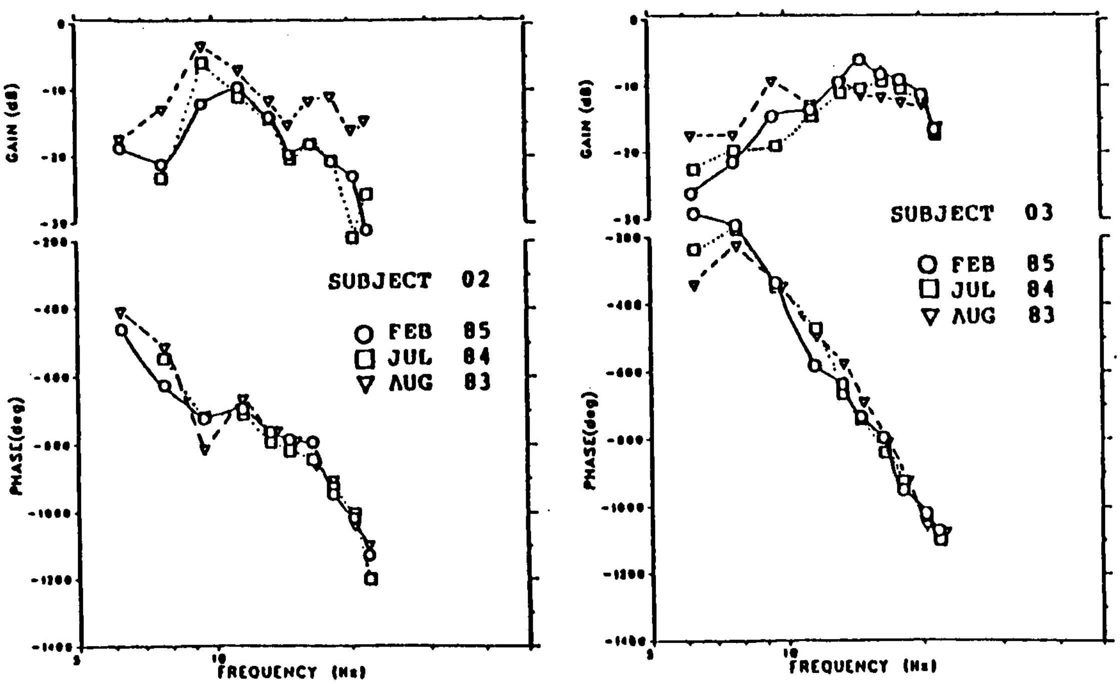


Figure 7. Repeatability of the SSEP as illustrated by describing function gain and phase values for 2 subjects over a 3 year span.

improvement of the visual-cortical response measure is to develop a closed-loop visual-cortical response paradigm. This requires providing an appropriate feedback signal to the subject.

From the evoked response data it was observed that evoked potentials could exhibit frequency responses as narrow as the measurement bandwidth of the experimental system being used, for example 0.0244 Hz (Junker et. al. 1987). Thus we concluded that frequency specificity of the feedback signal should be of concern.

If a feedback loop is to be effective it must also contain minimal transport delays. EEG biofeedback trainers at the Menninger Foundation (Biofeedback Center, Topeka, Kansas, personal communication) indicated that a biofeedback signal should not be delayed more than 4 cycles for it to be a useful signal from which subjects could learn to "control" their EEG.

From the above discussion, it was concluded that for the feedback signal to be effective it must be both timely and frequency specific. Useful feedback information about a 10 Hz response, for example, might require no more than a 0.4 second delay. To achieve this small delay and simultaneous frequency specificity is not an easy task. For the work reported above, a frequency specificity of 0.0244 Hz was

achieved, but only by analyzing 40.96 seconds of data at a time. Thus we concluded that frequency resolution and timeliness could not be achieved by our available digital apparatus. Instead, an analog active-filter approach was pursued.

The approach involved using a tunable bandpass filter in combination with a Lock-in Amplifier System (LAS). A diagram for this system is presented in Figure 8. The LAS consists of two quadrature phase sensitive detectors, the outputs of which are lowpass filtered and converted to polar form to yield continuous gain and phase signals at the lock-in frequency. The lock-in frequency is determined by a clock which generates a square wave, a quadrature square wave, and a sine wave. The square waves drive amplifiers A and B. The sine wave is used to drive the light stimulus. A narrow bandpass filter (tuned to the clock frequency) is used to improve the signal to noise ratio of the signal analyzed by the LAS. The responsiveness and frequency specificity of the LAS depends upon the cutoff frequency of the lowpass filters.

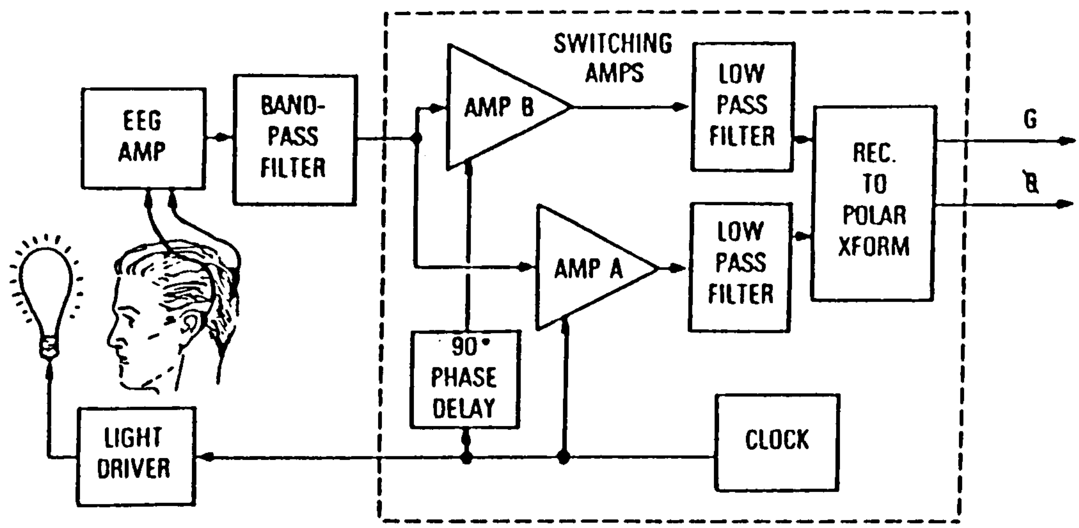


Figure 8. Lock-in amplifier system.

The LAS provides a continuous measure of gain and phase suggesting that it could be used in conjunction with steady-state stimulation to explore the time varying nature of task loading. A possible approach would be to stimulate with the SOS stimulus and continuously record the LAS output at one of the 10 SOS frequencies. Correlations between the continuous measure and the time varying nature of the task could be investigated. In the case of the decision making task this might be the times of appearance of new targets and times before or at the moment of button pushing.

The above is still an open loop measure. To close the loop using our approach, it was necessary to provide feedback to subjects of their EEG production at one or more evoking frequencies. The experimental setup we used to accomplish this is illustrated in Figure 9. Feedback of EEG production was provided to subjects through two modes: a light bar display, and an amplitude modulated tone. The qualifications for tone selection were that it be harmonically related to the evoking stimulus frequency and also subject verified as 'pleasing'. As the subject's EEG amplitude increased at the target frequency, as indicated by the LAS gain signal, more light bars became lit and the tone volume increased.

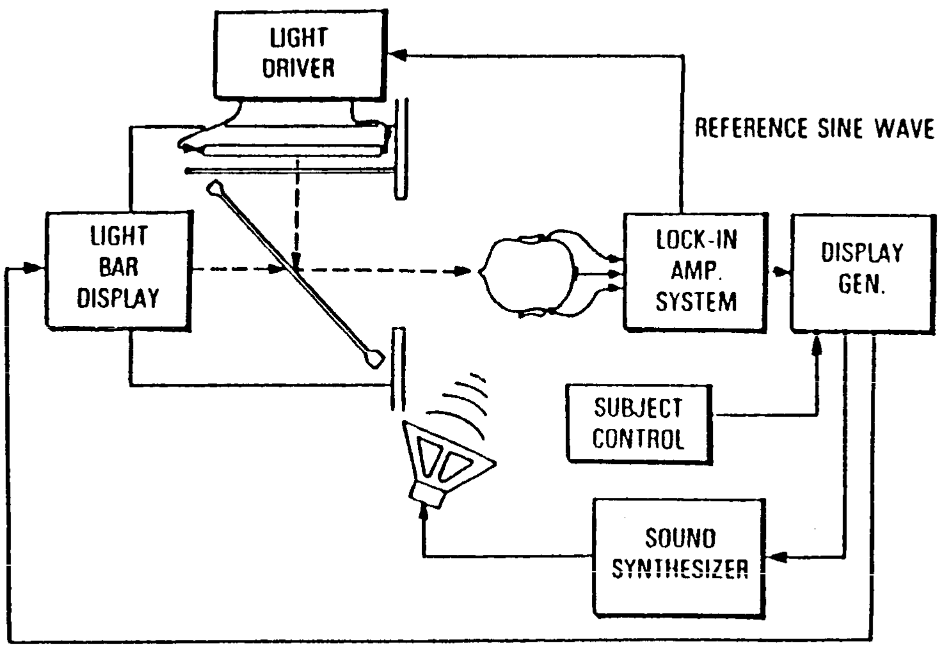


Figure 9. Experimental setup for feedback training.

For feedback training it was decided to use frequencies that would hopefully reside within relatively quiet areas of the EEG spectrum for the initial investigation. Therefore two frequencies were chosen, one below the alpha band and one between the alpha band and beta band. In addition the two frequencies were selected from the 10 sinewaves used in the SOS stimulus so that describing function data would be available for subsequent comparisons. Therefore frequencies of 7.73 Hz and 13.25 Hz were used.

To evaluate the effectiveness of feedback, two conditions were investigated. The first condition consisted of using the experimental setup as illustrated in Figure 9. One group of subjects trained under this condition. For the

second condition, true EEG feedback was replaced by false feedback from an analog random noise generator. This output was injected into the bandpass filter of the experimental setup instead of the subject's EEG (refer to Figure 8). A second set of subjects was used for this false feedback condition. The subjects, although aware of the possibility of getting either real or false feedback, were not informed until the experiment's conclusion as to which type of feedback they had received. After receiving 6 sessions of false feedback these subjects received true feedback for 4 sessions.

The four subjects used for the decision making task investigation (Figure 5) were used in this experiment. Subjects were randomly assigned to the two experimental groups with the constraint that the two alpha producers (Subjects 13 and 77) would not be in the same group. This resulted in Subjects 13 and 07 being assigned to the true feedback group and Subjects 77 and 03 to the false feedback group.

To provide comparable results between subjects for each frequency under investigation, the EEG response was adjusted to approximately the same level for each subject at the start of each session. A variable gain control of the EEG signal prior to the bandpass filter (refer to Figure 8) was used to achieve EEG gain adjustment. The result of this adjustment was determined by monitoring the subject's EEG spectrum with an HP Fourier analyzer at the output of the variable gain control.

For each experimental session, subjects trained at both frequencies. The first half of the session consisted of training at one frequency and the next half at the second frequency. The task of the subject was to either increase the feedback signal or decrease the feedback signal over a 100 second trial. An experimental session consisted of two blocks of eight 100 sec periods for each frequency or a total of 4 blocks per session. Within each block of 8 trials, subjects were instructed to "raise the light bar" (increase the feedback signal) for 4 trials, and "lower the light bar" (decrease the signal) for 4 trials. The order of presentation of the two frequencies as well as the order of raising and lowering was randomized.

One mode of EEG control is the ability, at a given frequency, to hold one's amplitude above or maintain it below a hypothetical threshold. The fifth light bar on a 16 light bar display was chosen as a threshold. Performance scoring was a measure of how many seconds, out of a 100 second trial, the subject's amplitude went above this fifth bar level. The second performance measure was the coherence between subject EEG and the evoking light stimulus. For each block of eight trials, the average difference for each performance measure

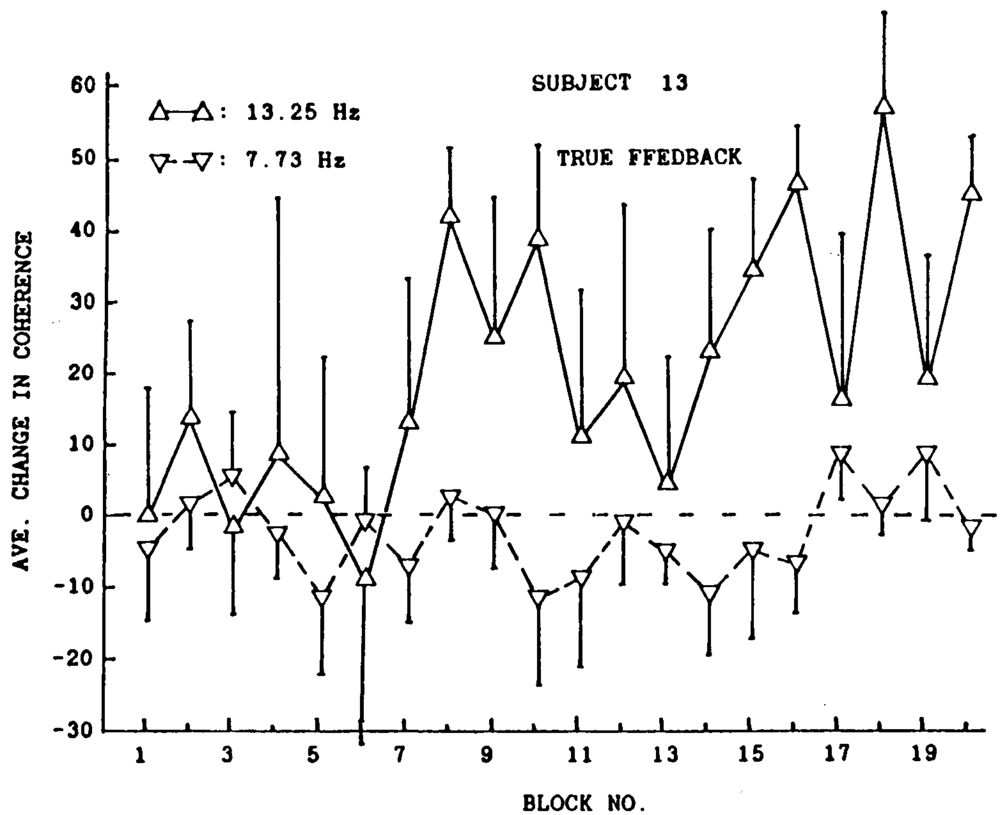
between increasing and suppressing the EEG signal was computed. This resulted in average performance scores and standard deviations for both increasing and suppressing EEG signals for each block. The results of this analysis are presented in Figures 10 and 11. Plotted in each graph are the average values and the largest standard deviation (either from increasing or suppressing) per block. A value above the dashed line in each graph indicates for that block the average of the 4 'increasing' values was greater than the average of the 4 'suppressing' values. Values below the dashed line indicate that the opposite trend occurred.

DISCUSSION OF FEEDBACK TRAINING RESULTS

Before beginning discussion of the feedback training results it is informative to refer to the Subjects' describing functions and remnant spectra of Figure 5. Looking first at Subject 13's responses, a weak response at the lower frequency (7.73 Hz) as indicated by the large standard error bars for the three conditions tested can be observed. The response at 13.25 Hz, compared to the alpha response at 11.49 Hz for the lights only condition, was low but increased with task loading. Subject 77's responses at both frequencies were low and weak as indicated by the mean values and the large standard error bars. Subject 07 exhibited large variability in the evoked response at 7.73 Hz. Subject 03's response at 13.25 Hz for the lights only condition was weak.

The coherence results for Subject 13 at 7.73 Hz (Figure 10a) indicate that no net change in coherence occurred due to feedback training. Over the 20 blocks, the average value in coherence was only slightly greater when suppressing than when increasing. At 13.25 Hz, however, by the seventh block a consistent increase in coherence between the increasing and suppressing trials can be observed. The lack of change in coherence at 7.73 Hz may relate to the weak response obtained in the Subject's describing functions of Figure 5a. Subject 07 exhibited similar trends in both the average change in coherence and in the describing functions of Figure 5b.

Data for the subjects receiving false feedback for 6 sessions (12 blocks) and then true feedback for 4 sessions are shown in the second two graphs of Figure 10. Subject 77 exhibited greater average coherence during the increasing trials for 13.25 Hz, even during the false feedback conditions. Due to the large variation in the data however this trend was not very consistent. Subject 03 exhibited greater coherence during the increase trials as compared to the suppress trials at 7.73 Hz, but not at 13.25 Hz. This corresponds to the gain sensitivity observed for Subject 03 in Figure 5b.



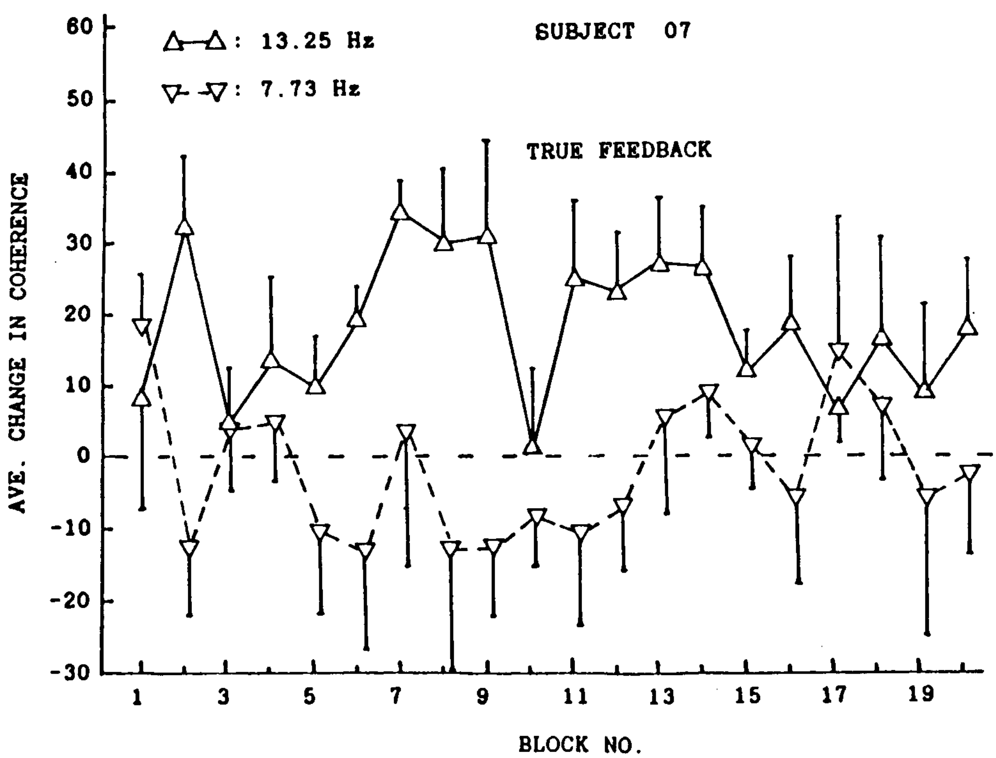
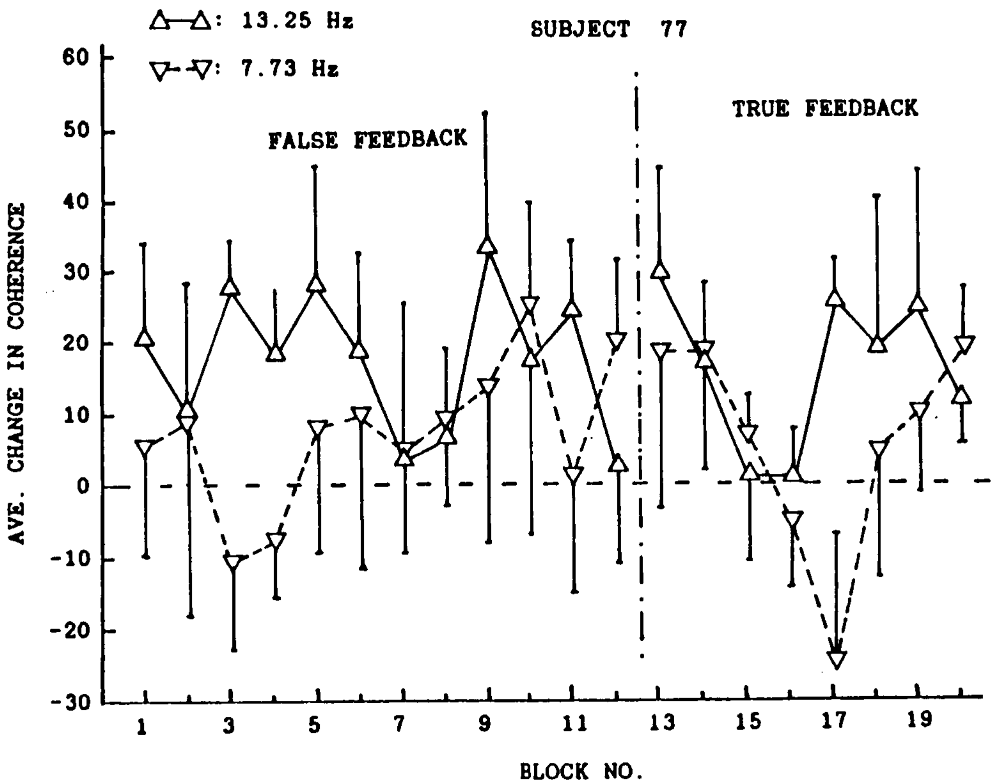


Figure 10a. Average change in coherence for subjects with true feedback, standard deviation bars included.



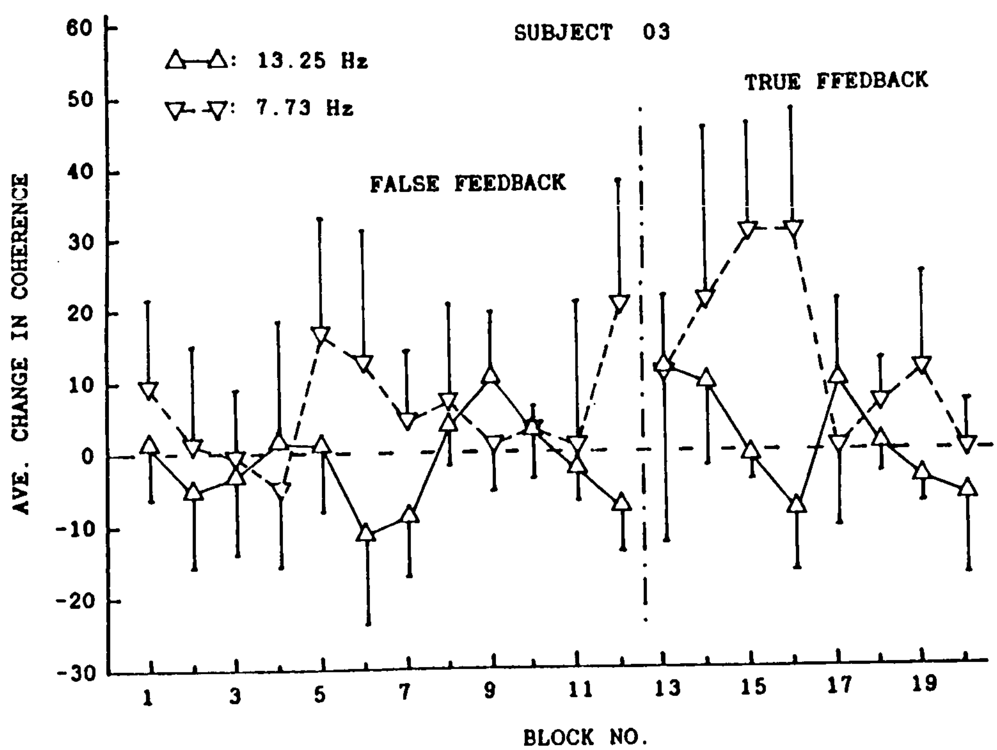
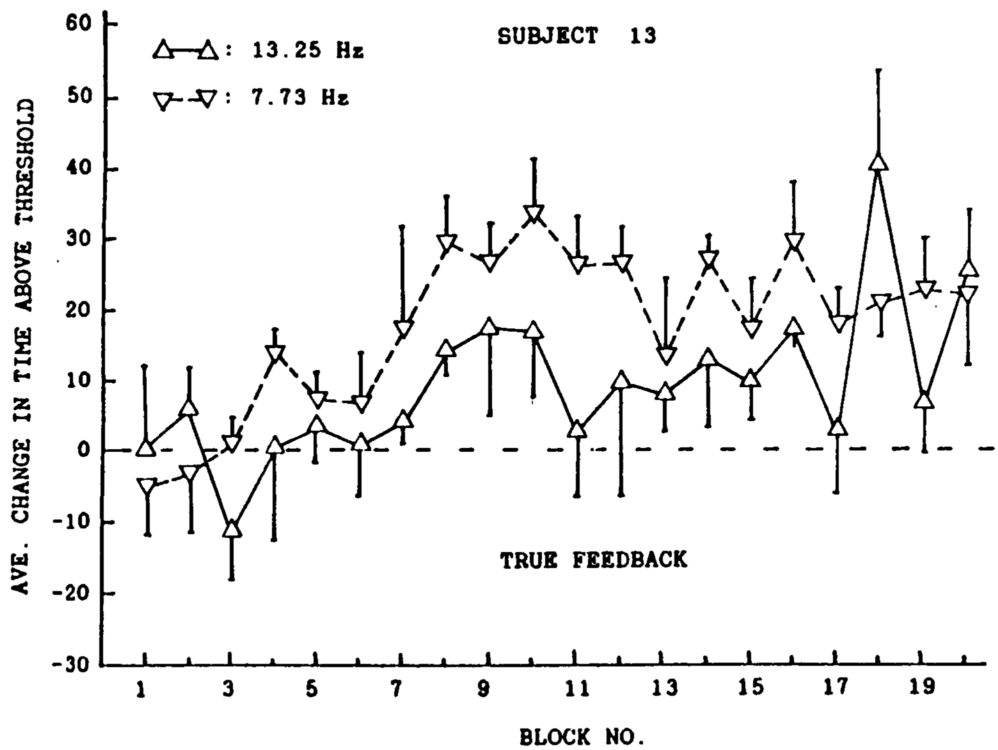


Figure 10b. Average change in coherence for subjects who received false feedback for the first 12 blocks.



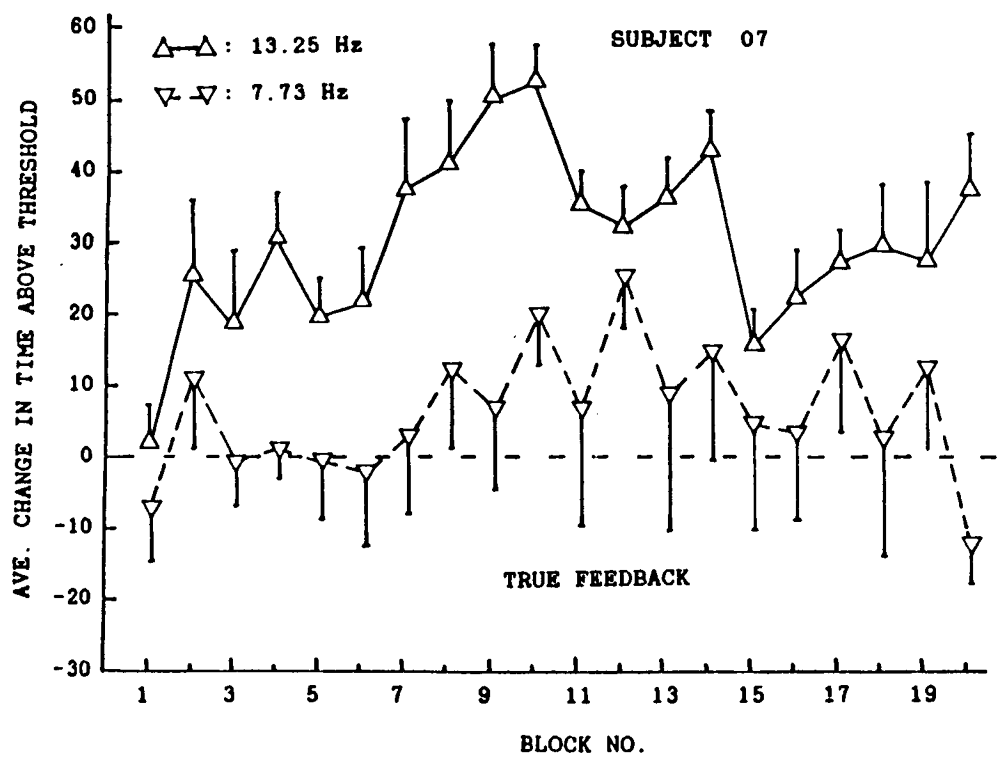


Figure 11a. Average change in time above threshold for subjects who received true feedback. Positive values indicate longer time above threshold for increasing trials than for suppressing trials.

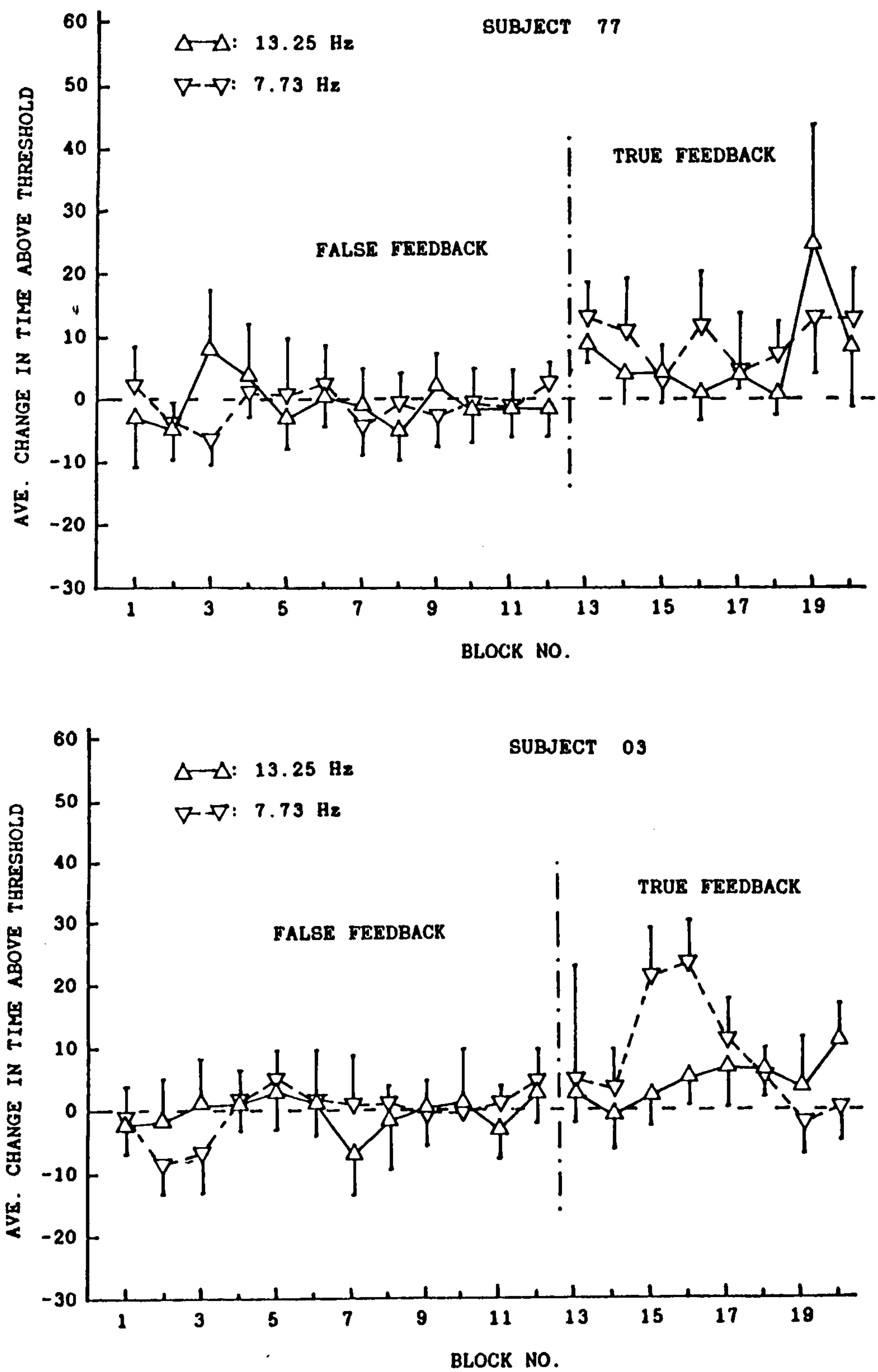


Figure 11b. Average change in time above threshold scores for subjects receiving false feedback for 12 blocks, and then true feedback for 8 blocks.

For the true feedback group, consistent positive trends in coherence were exhibited by subjects only after 7 blocks. Since the false feedback group only had 8 blocks of true feedback training, it is not unexpected that no conclusive trends in coherence were observed.

In contrast to the coherence results previously discussed for Subject 13, this Subject's positive average change in time above threshold (Figure 11a) was consistently higher for 7.73 Hz than for 13.25 Hz. Note that it took at least 4 sessions (8th block) before consistent control began to occur. Blocks 18 and 20 indicate that a big step in learning at 13.25 Hz had occurred. Subject 07 exhibited strong consistent control at 13.25 Hz and marginal control at 7.73 Hz.

For the second group, during the false feedback trials, as to be expected the average time above threshold was approximately zero as it was a result of noise. The plots for Subjects 77 and 03 during false feedback are actually plots of what they saw and heard in terms of feedback cues. When given true feedback both subjects began to exhibit positive average times above threshold indicating EEG control. With further sessions improvements similar to those observed for Subjects 13 and 07 might be expected.

CONCLUDING REMARKS

From the results of Figure 11, it can be concluded that conscious control of EEG at specific frequencies corresponding to evoking stimuli can be achieved. Further, this conscious control can affect the coherence of the response. This has interesting implications relative to the question of the appropriateness of using the SSEP for mental-state estimation. The subject's ability to manipulate their EEG levels is continually and unpredictably active and without the harnessing effects of feedback it may alter SSEPs in an unforeseeable manner. Thus open loop measures may be fraught with uncontrollable changes. A possible solution would be to employ the feedback paradigm reported here during performance so that subjects could be kept continuously aware of their mental state.

As configured in Figure 8, the LAS may be too slow in responding or not sufficiently frequency specific to provide the most effective feedback signal. For large amplitude or large phase variations in the EEG at the reference frequency this will be true. For small perturbations, once a feedback loop has been achieved, LAS response time may be acceptable.

Extending the lowpass filters' cutoff frequencies improves the LAS response time but increases the bandwidth. A possible improvement to the LAS may be the addition of a

phase-locked loop. In a typical phase-locked loop system the reference frequency is made to follow the phase of the incoming signal for stability. Utilizing analog delay lines to shift the phase of the reference sine wave as it drives the light stimulus may achieve the desired effect. The approach would be to delay the sine wave one complete cycle and lead or lag an additional amount, determined by the phase signal of the LAS. The intention of this approach would be to provide a more effective evoking stimulus so that the visual-cortical system knows it is "looking at itself."

In closing, it has been shown that with appropriate loop closure humans can achieve narrow-band frequency control of their brain waves. This ability leads directly to control of brain actuated systems. Furthermore, two humans actuating the same control may be the foundation of brain-to-brain communication.

Considering the neurophysiology of the brain near the surface (Guyton, 1986), the cortex is rich in dendritic connections. This evokes the image of a sensitive radio receiver/transmitter. Perhaps in the future the equipment and technology discussed will not be needed to achieve brain actuated control and brain-to-brain communication. At this time, however, the technology presented can help to open the way, while providing insight into the workings of the human brain and a handle on mental-state estimation.

REFERENCES

Guyton, A.C. 1986. Textbook of Medical Physiology. 5th Edition, W.B. Sanders Co.,ISBN 0-7216-4393-0.

Junker, A.M., Levison, W.H., and Gill, R.T. 1987. A Systems Engineering Based Methodology For Analyzing Human Electrocortical Responses. AAMRL-TR-87-030, Armstrong Aerospace Medical Research Laboratories, WPAFB, Ohio, 1987.

Levison, W.H. 1983. Development of a Model for Human Operator Learning in Continuous Estimation and Control Tasks. AFAMRL-TR-83-088, Air Force Aerospace Medical Research Laboratory, WPAFB, Ohio, December 1983.

Levison, W.H., Elkind, J.I., and Ward, J.L. 1971. Studies of Multivariable Manual Control Systems: A Model for Task Interference. NASA CR-1746, May 1971.

Levison, W.H., and Junker, A.M. 1978. Use of Tilt Cue in a Simulated Heading Tracking Task. Proceedings of the Fourteenth Annual Conference on Manual Control, Los Angeles,Ca.

Pattipati, K.R., Ephraph, A.R., and Kleinman, D.L. 1979. Analysis of Human Decision-Making in Multitask Environments. University of Connecticut, Technical Report EECS-TR-79-15.

Regan, D. 1972. Evoked Potentials in Psychology, Sensory Physiology and Clinical Medicine. Chapman and Hall Ltd., London, SBN 412-100920-4, 1972.

Shingledecker, C.A., Acton, W.H. and Crabtree, M.S. 1983. Development and Application of a Critereon Task Set for Workload Metric Evaluation. Proc. of 2nd Ann. Aerospace Behavioral Engineering Technical Conf. Aerospace Congress and Exposition Press. Long Beach, Ca.

Spekreijse, H. 1966. Analysis of EEG Response in Man Evoked by Sine Wave Modulated Light. Dr. W. Junk Publishers. The Hague, Netherlands. 152 pp.

Wilson, G.F. and O'Donnell, R.D. 1980. Human Sensitivity to High Frequency Sine Wave and Pulsed Light Stimulation as Measured by the Steady State Cortical Evoked Response. Technical Report AFAMRL-TR-80-133.

Zacharias, G.L. and Levison, W.H. 1979. A Performance Analyzer for Identifying Changes in Human Operator Tracking Strategies. AMRL-TR-79-17, Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio, March 1979.

N88-23381 S11-53 140799 8P. N0052345 BG535384

VOICE-STRESS MEASURE OF MENTAL WORKLOAD

Murray Alpert
New York University
New York, NY

Sid J. Schneider
Behavioral Health Systems, Inc.
Ossining, NY

In the 1970's, several studies employed voice analysis as a measure of workload. These studies usually looked at the suppression of the 8 to 12 Hertz microtremor in human voice as a measure of stress (Ref. 1). The existence and significance of the microtremor was controversial and the initial interest waned. Since then, a number of approaches have been developed, directed toward a detailed and extensive analysis of speech prosody. This method is intuitively appealing, since the emphasis, rhythm and inflection of a person's voice would seem to reflect psychological variables like stress, affect and the demands being placed on the speaker.

Research which explores the relationship between speech prosody and workload is relevant to the advanced flight deck. Flight crews will be making increased use of voice technology; the advanced flight deck will "speak" using voice synthesis and receive commands verbally from the crew. Therefore, speech samples will be readily available from flight crews in advanced flight decks. An apparatus that could assess the mental state of the flight crew from voice samples could be useful in the design and evaluation of the advanced flight deck.

The apparatus to be described was originally designed for applications in psychiatry, to provide objective and quantitative measures of variations in feeling states such as depression, mania, or the flat effect of schizophrenia (Ref. 2). Also of interest was the measurement of medication effects on such psychiatric states. Empirical studies have targeted the most discriminating acoustic parameters for each of these variables. Similarly, empirical studies could identify those acoustic measures which are most sensitive to the effects of variations in workload in aircraft situations.

This hybrid analog/digital analyzer provides information about such basic speech variables as fundamental frequency (pitch), amplitude (loudness), the duration of utterances and pauses, and the variances of these measures. The system consists of three main components: 1) a good quality stereo cassette tape deck; 2) a microcomputer equipped with an analog to digital conversion system (Northstar Horizon with a Tecmar TM-AD212 analog to digital converter); and 3) a multifunction analog signal processing unit. The analog computer unit provides the circuitry for filtering and transforming the speech signals prior to digital analysis. The raw AC signal that comes from the tape deck is first passed through a bandpass filter in order to restrict the signal to a range around the speaker's fundamental frequency, and to eliminate harmonic frequencies. The range between the filter can be adjusted for the particular voice; for example, it is usually set between 80 to 100 hertz for male voices and between 120 and 300 hertz for female voices.

Once filtered, the speaker's signal is then split into two parallel lines

which are analyzed separately, one channel for frequency information and one for amplitude information. The frequency signal goes through a frequency to voltage converter which outputs 1 volt for each 50 hertz of signal; this signal then goes to one of the channels on the A/D converter board with a resolution of 200 counts per volt. The resulting resolution is 4 counts per hertz. The signal on the amplitude line is first passed through an attenuator, then full wave rectified to a DC signal and finally demodulated to produce a smooth signal that goes to another channel on the A/D converter board.

In the software there is a log lookup table so that the variation in voltage and frequency across time is made proportional to the logarithm of the amplitude and frequency of the voice. An utterance is defined as an amplitude which is above some threshold of background noise for 100 msecs or more; a gap as an amplitude that goes below threshold for at least 200 msecs; and a peak as a point of maximum amplitude relative to the values of amplitude immediately preceding and following that point.

The software was designed to measure the following prosodic features of speech:

1) Number of utterance, gaps, and peaks.

2) Mean and variance of the time durations of utterances, gaps, and peaks.

3) Mean and variance of the natural log of the amplitude of peaks as well as the log of the frequencies corresponding to those peaks.

4) The correlation between peak amplitude and peak frequency.

5) The distribution of peaks within utterances (i.e., how many 1 peak utterances, 2 peak utterances, etc. were there) as well as summary information about the duration of the peaks in those utterances.

The hardware and software allow for the setting of a threshold to eliminate background noise. It is also possible to remove the effects of other speakers. Their speech, recorded on the second channel of the stereo deck, can be sent to a separate channel of the analog computer, which detects the presence of a signal and sends a TTL signal, detected by the software, which suspends the analysis until the TTL signal is removed. In this way, the speech that is analyzed is uncontaminated by other speakers and noise that may be present in an aircrew operational setting. A calibration signal of known amplitude and frequency is recorded on the subject's channel. Since the subject uses a head mounted microphone of known output, the use of a calibration signal permits a usable estimate of the absolute voice level.

Results of Previous Studies

The apparatus for analysis of the human voice has been employed in a series of clinical studies at the Millhauser Laboratories for Research in Psychiatry and the Behavioral Sciences at New York University Medical Center. Several studies have suggested that data from the apparatus are reproducible, highly precise, and useful in a clinical setting. For example, the apparatus provides an objective, reliable means of quantifying flat affect--the restricted emotions apparent in many schizophrenics--and distinguishing it from the clinically very similar presentation of patients with a retarded depression (Ref. 3). Flat affect is diagnostically important in schizophrenia. However, it is difficult to measure because other processes,

such as psychomotor retardation, institutionalization, and drug side effects can mask it. Voice analysis provides a way to quantify flat affect in schizophrenics on the basis of diminished inflection (variation in frequency) and diminished dynamics (variation in volume). In depressives, the mood disturbance tends to be shown by long pauses and brief utterances (Ref. 4). Measurement of flat affect using this apparatus compared favorably to clinical ratings made by highly skilled attending psychiatrists in evaluating and predicting patient behaviors (Ref. 5).

Acoustic analysis has permitted the articulation of processes that are frequently confounded clinically and conceptually. Thus, it has become possible to distinguish effects from moods. Affects are encoded in voice emphasis. Affects are visible as the rapid fluctuations in the acoustic parameters amplitude and number of multi-peak utterances. Affects, such as excited affect, reflect momentary feelings of which the speaker may not be entirely conscious. Moods, on the other hand, are encoded in temporal patterns of utterances and pauses and have much slower temporal phases. Moods are the subjective feelings, like sadness and joy. They are revealed in the length of pauses and utterances. It is important to distinguish both of these processes from emotions, like anger or fear, which are detectable in voice because they disrupt normal speech patterns. If a subject is emotionally aroused, the arousal affects physiological mechanisms important for speech. Changes in respiration will affect speech energetics; changes in muscle tone will alter the overtone structure and the speaker's voice quality (Ref. 6). These changes are visible as alterations in voice frequency lasting several minutes.

These insights into the separation of different feeling states grew out of studies with a variety of patient populations, treatment paradigms, and experimental procedures for producing emotional arousal, such as having the subject lie or by applying mildly aversive stimuli. It is noteworthy that these procedures can produce a vocal broadcasting of emotional arousal in patients with depression or schizophrenia as well as in controls. The different feeling states appear to be controlled by different and perhaps orthogonal brain mechanisms.

The apparatus has not been applied to the study of man-machine interactions. However, many of the psychological variables of interest in a clinical setting, like attention, arousal, and affect would also be of interest in human factors studies. The approach may well be appropriate to the study of multidimensional variables like workload. We have begun a study to determine which features of voice prosody reflect the workload experienced by the speaker. As of this writing, only two preliminary subjects have been run, but their results can be reported.

SUMMARY OF PROCEDURE

Subjects will be males between 18 and 50 years of age without uncorrected sight, speech, or hearing condition. Subjects will be run individually in a windowless room free of distractions. The subject will be seated at a Taxan 630 computer screen and have before him a hand held momentary contact switch. He will wear a set of headphones attached to a Shure head mounted microphone. White noise at 60 dB (0.0002 microbar reference) will be presented over the headphones to simulate cockpit noise.

An IBM PC AT computer will present simultaneous primary and secondary tasks. The primary task will be designed to be simple enough to be performed errorlessly. It will require speech and will be the source of the voice samples used for analysis. The secondary task will be used to manipulate workload. The system will continually monitor the error rate in the secondary task and adjust the presentation rate in order to keep the error rate constant. There will be a high workload (i.e., high error rate) and low workload (i.e., low error rate) condition.

In the secondary task, the numerals 1 through 6 will be presented in a predetermined order, one after the other, in the center of the screen. The subject's task is to press the button immediately whenever two numbers in a row total seven. Thirty percent of the numerals will be targets. While these numbers are presented, there are two triangles, one on either side of the central numbers, about 7 cm away. At intervals ranging from 18 to 28 sec, one or the other triangles, randomly chosen, will appear to rotate. The subject must state, as rapidly as possible after the triangle begins to rotate, "The triangle that started moving should stop now." The triangle will in fact stop rotating upon voice offset (or after 10 seconds pass). This speaking task is the primary task.

The computer will automatically record the number of correct responses in the secondary (number) task, as well as the reaction times of the correct responses. The number of commission, omission, double strike and late errors will also be recorded. In order to minimize the effect of speaking itself, the system will not record performance on the secondary (number) task while the subject is speaking as part of the primary task.

The reaction time of the voice in the primary task will be recorded. The voice itself will be captured on a cassette deck for analysis on the voice prosody analytic apparatus at the Millhauser Laboratories, New York University School of Medicine.

The session for each subject will begin with a series of short practice trials designed to familiarize the subject with the tasks. The trials will also suggest which presentation rates for the central number task result in error rates of 20 and 60 percent. These rates will be the initial presentation rates used, in the low and high workload conditions, respectively. The system is programmed to adjust the presentation rate at predetermined intervals to maintain the error rate at the desired level.

The low and high workload tasks will be presented in 10 minute segments. For half of the subjects, the order will be LHHL, and for the other half, HLLH.

Analyses of the cassettes should reveal which aspects of voice prosody are associated with increased workloads. The error rates and reaction times in the central number task will corroborate the assertion that mental loading has in fact been manipulated by changing the rate of number presentation. A faster presentation rate should increase error rate and decrease reaction time. The continual adjustment of the presentation rate will insure that workload transients are avoided to the extent possible.

Preliminary Results

Two preliminary subjects have been run in this study. Both were female. Their data will not be used in the final report of this research. Table 1 shows the error rates for the secondary (number) task in each of the two 10 minute runs in the high and low workload conditions. The table reveals that the software was effective in maintaining error rates close to the intended error rates. Table 2 shows the average reaction times for the button pushes in the correct responses in the number task. The high workload condition, in which the presentation of the numbers was fast, brought about faster reaction times than the low workload condition did. Table 3 shows that the voice reaction times (the length of time between the moment that the triangle started turning and the subject spoke) also tended to be faster in the high workload condition. The standard deviations of the voice and number tack reaction times, shown in parentheses in the tables, tended to be higher in the high workload condition, as compared with the low workload condition. These results would suggest that the speeded presentation of the numbers in the high workload condition was in fact successful in bringing about increased workload.

Table 4 shows the results of the analysis of the voice of the two subjects. The table reveals a trend for the frequency and amplitude of the voice to increase with each successive run, regardless of whether the run was in the high or low workload condition. However, these preliminary results suggest a possible trend for the voice frequency and amplitude to be higher, and the variance of the voice frequency to be lower, in the high workload condition. These results would replicate a previous study (ref. 6). However, this study must be run with the sample of 15 subjects before any conclusions can be drawn. Also, in this preliminary study, 25 voice samples were obtained in each run. This number of samples made the runs rather long and aversive to the subjects. In the actual study, the runs will be shortened to 15 samples. The means for the first 15 samples from the preliminary subjects were similar to the means for all 25 samples. By running a larger number of subjects in shorter runs, the effect of workload upon voice prosody should become more apparent.

REFERENCES

1. Schiflett, S.G. & Loikith, B.S.: Voice stress analysis as a measure of operator workload. Naval Air Test Center, Patuxent River, Maryland, 1980.

2. Alpert, M., Homel, P., Merewether, F., Martz, J. & Lomask, M.: Voxcom: A system for real time analysis of natural speech. Presented to the Eastern Psychological Association, New York City, 1986.

3. Mayer, M., Alpert, M., Stastny, P., et al.: Multiple contributions to clinical presentation of flat affect in schizophrenia. Schizophrenia Bulletin, vol. 11, 1985, pp.420-426.

4. Alpert, M.: Feedback effects of audition and vocal effort on intensity of voice. Journal of the Acoustical Society of America, vol. 39, 1966, pp. 1218.

5. Alpert, M. & Anderson, L.T.: Imagery mediation of vocal emphasis in flat

affect. Archives of General Psychiatry, vol. 124, 1985, pp. 202-211.

6. Shipp, T., Brenner, M., & Doherty, E. T.: Vocal indicators of psychophysiological stress induced by task loading. USAF SAM-TSQ-86-3, September, 1986.

ERROR RATES

	High Workload		Low Workload	
	First Run	Second Run	First Run	Second Run
DESIRED	.60	.60	.20	.20
SUBJECT 1	.64	.61	.19	.20
SUBJECT 2	.60	.62	.21	.20

Table 1. The desired error rate in the secondary (number) task in the high workload condition was .6; in the low workload condition, the desired error rate was .2. Error rate was defined as (number of errors) / (number of errors + number of correct responses). The software was able to maintain subjects' performance near the desired error rates.

NUMBER TASK REACTION TIME (msec)

	High Workload		Low Workload	
	First Run	Second Run	First Run	Second Run
SUBJECT 1	312 (204)	335 (194)	586 (126)	538 (127)
SUBJECT 2	329 (230)	353 (210)	663 (157)	652 (164)

Table 2. Entries are the mean reaction times in msec for correct responses in the secondary (number) task. The reaction time was defined as the time between the appearance of a target number (the second number of a pair that added to 7) and the moment the subject pressed the switch. Standard deviations are in parentheses. The high workload condition appears to have brought about faster reaction times and higher standard deviations.

VOICE REACTION TIME (msec)

	High Workload		Low Workload	
	First Run	Second Run	First Run	Second Run
SUBJECT 1	804 (104)	716 (114)	820 (79)	808 (83)
SUBJECT 2	988 (400)	1030 (263)	998 (213)	950 (191)

Table 3. Entries are the mean reaction times in msec for the primary (voice) task. The reaction time was defined as the time between the initiation of triangle movement and speech onset. Standard deviations are in parentheses. The high workload condition may have brought about faster reaction times and higher standard deviations.

Acoustic Measures

(summary data - 25 sentences per run)

	Work-load	Run #	Uttdur	Uttvar	Mean Amp	Var Amp	Mean Freq	Var Freq
SUBJECT 1	low	1	184.5	14.5	458.3	249.7	520.4	42.9
	high	2	185.2	2392.6	469.8	223.6	523.9	28.5
	high	3	211.6	5925.7	493.6	164.2	531.3	10.3
	low	4	205.2	3662.0	502.5	164.7	530.0	10.5
SUBJECT 2	high	1	180.4	3954.3	435.7	149.3	521.8	23.0
	low	2	180.4	1374.1	439.6	136.7	524.1	6.3
	low	3	190.6	126.3	437.1	202.3	521.5	7.4
	high	4	207.1	5336.6	448.4	169.7	528.4	5.6

N88-23382

5/2-53
140800
16 P.

PRIMARY TASK EVENT-RELATED POTENTIALS RELATED TO DIFFERENT ASPECTS OF INFORMATION PROCESSING[1]

Robert C. Munson, Richard L. Horst, and David L. Mahaffey
ARD Corporation
Columbia, MD

AD 234656

ABSTRACT

This paper reviews the results of two studies which investigated the relationships between cognitive processing and components of transient event-related potentials (ERPs) in a task in which mental workload was manipulated. The task involved the monitoring of an array of discrete readouts for values that went "out-of-bounds," and was somewhat analogous to tasks performed in cockpits. The ERPs elicited by the changing readouts varied with the number of readouts being monitored, the number of monitored readouts that were close to going out-of-bounds, and whether or not the change took a monitored readout out-of-bounds. Moreover, different regions of the waveform differentially reflected these effects. The results confirm the sensitivity of scalp-recorded ERPs to the cognitive processes affected by mental workload and suggest the possibility of extracting useful ERP indices of primary task performance in a wide range of man-machine settings.

INTRODUCTION

There is by now a vast literature relating scalp-recorded brain electrical activity to various cognitive processes. Other talks in this session have focused on studies which related behavioral performance to either steady-state evoked potentials, elicited by a rapidly oscillating stimulus, or probe evoked potentials, elicited by discrete stimuli that were irrelevant to the mental processing task the subject was given. In contrast, the data presented in this paper relate to the use of the transient evoked potential (or event-related potential, i.e. ERP) elicited by task-relevant stimuli. In particular, we examined the scalp-recorded responses to discrete visual stimuli that were presented in the context of a monitoring task as the mental workload of that task was systematically manipulated.

Most previous investigations that have addressed the relationship between ERPs and mental workload have focused on responses elicited in dual-task paradigms. Typically the waveshape of the ERP elicited by secondary task stimuli has been related to changing levels of difficulty of the primary task and has been interpreted as reflecting the spare cognitive capacity that remains after the demands of the primary task have been met. While the results of these studies have revealed important insights regarding the influence of cognitive processes on ERPs, it is not clear how widely applicable this methodology will be in evaluating the workload of human operators in real-world

---

[1]The research described herein was funded by NASA under Small Business Innovative Research Contracts to ARD Corporation (NAS 1-17576, 1-18019). Dr. Alan Pope was the contract technical monitor. The studies were conducted in Dr. Daniel Ruchkin's laboratory at the University of Maryland Medical School. Dr. Ruchkin provided insightful guidance in data analysis. Ms. Anne Francoeur provided valuable assistance in data collection and data processing.

systems.

The secondary tasks used in most laboratory studies of mental workload have been relatively simplistic and contrived. They have been chosen for the convenience with which their stimuli elicit the responses of interest, whether physiological or behavioral. Although such tasks offer a conceptual similarity to operational systems in which human operators must time-share between tasks and process stimuli which compete for attention, they do not lend themselves readily to use in operational or simulated systems. In most operational systems in which mental workload is a concern, the operator is already over-burdened. To further burden him with contrived stimuli and tasks, in order to assess the workload of existing tasks, is impractical at best and invalid at worst. Even if certain existing tasks offer stimuli to which ERPs, and reaction times can be time-locked, it is unlikely that they will be functionally equivalent to the contrived secondary tasks used in the laboratory. Such "secondary" tasks will likely be performed in conjuction with differing configurations of other existing tasks, and it is difficult to ensure that these other "primary" tasks are given priority, as implicitly assumed if one is to interpret secondary task measures as reflecting spare cognitive capacity.

Because of these considerations, we examined ERPs that were elicited by stimuli presented in a single (primary) task as the difficulty of that task was varied. Workload-related effects obtained in such a paradigm would suggest the usefulness of ERP measures of cognition, both for systems in which processing resources can be devoted to a single task, as well as those in which the ERP-eliciting task must be time-shared with others. The present paper summarizes the results of two studies for which more detailed accounts of methods and results have already been published (refs 1, 2, 3). Our intent here is to discuss these studies, both the reasoning behind them and the interpretation of results, in terms of their implications for eventual applications in man-machine systems.

WORKLOAD EFFECTS ON TRANSIENT ERPS

Transient ERPs are usually extracted from the ongoing EEG by signal averaging over numerous occurrences of the eliciting stimulus. The ERP waveform is comprised of various "components", each having a characteristic scalp topography, latency range, and polarity. It is assumed that these components reflect the electrical activity from numerous generators within the brain, the activity of which overlaps in both space and time. For our purposes, it is not critical to understand the brain loci and generator mechanisms underlying these scalp-recorded components. Instead the focus is on how these components vary differentially with experimental manipulations and what these systematic variations suggest about the mental operations that these manipulations call into play. The components of most interest here are those which have been shown by previous work to be related, not to the physical characteristics of the stimuli to which an ERP was time-locked, but to the cognitive processing which was required by the task within which these stimuli were presented. Differential scalp topography and differential response to manipulations of the cognitive task are the primary means for disentangling the functional components of these waveforms.

Studies relating ERP components to mental workload grew out of previous findings which showed consistent attention-related effects on the amplitude of

the P300 component. P300s are elicited by stimuli that are attended (i.e. task relevant) and, in some sense, unpredictable (see e.g., reviews in refs 4, 5). The basic hypothesis underlying most studies of P300 and workload has been that P300 amplitude would be modulated by the amount of attention, or the amount of central processing resources, that could be devoted to processing the ERP-eliciting stimuli. Thus, in dual-task situations, when the attentional demands of the primary task are increased, there is less of the limited pool of attention that can be devoted to secondary task stimuli, and hence the amplitude of the P300 elicited by such secondary task stimuli should decrease. Much of this work has been performed by Donchin, Wickens, and their colleagues, at the University of Illinois (see review in ref. 5). In the early studies, tracking a computer-driven cursor was used as the primary task. The secondary task involved the presentation of discrete stimuli which required either an overt response to which choice reaction time was measured, or a covert updating of a running count of the occurrence of some subset of the stimuli.

The initial results were somewhat discouraging. The amplitude of the P300 elicited by low probability auditory stimuli in a counting task was markedly reduced when the counting was performed concurrently with a visual-motor tracking task; however, there were no further systematic decreases in the P300 amplitude as the difficulty of the tracking task was increased, either by requiring that tracking be performed in two dimensions (ref. 6) or by increasing the bandwidth of the cursor in a one-dimensional tracking task (ref. 7).

More encouraging results were obtained when the auditory counting task was time-shared with a visual monitoring task in which subjects detected directional changes in a simulated air traffic control display. In this situation, the P300 elicited by auditory stimuli decreased in amplitude as a function of the number of elements which subjects monitored (ref. 8). The interpretation of these findings was consistent with the viewpoint which was emerging from behavioral studies at the time (e.g., ref. 9) which posited that processing resources were segregated into multiple "pools." Thus P300 amplitude elicited by secondary task stimuli may have been modulated by the demands of the primary task when it involved visual monitoring, because the perceptual demands of these two tasks may have tapped the same pool of processing resources. On the other hand, the P300 amplitude elicited by secondary task auditory stimuli may not have reflected the workload dynamics of the tracking tasks, because the visual-motor demands of tracking tapped a different pool of resources.

Further evidence that P300 amplitude is related to available processing resources was sought by examining the reciprocity between the amplitudes of the P300 elicited in the context of primary versus secondary task stimuli in dual task paradigms. In order to elicit ERPs related to primary task processing, a task was developed which involved compensatory tracking with the cursor moving in discrete steps, rather than moving continuously as before. When subjects tracked these step changes in conjunction with a secondary task that consisted of counting occurrences of certain auditory stimuli, the amplitude of the P300 elicited by the secondary task stimuli decreased as the difficulty of the tracking task increased. However, when subjects were instructed to count occurrences of the cursor step changes in a given direction (i.e., the secondary task stimuli were "embedded" in the primary task), the P300 elicited by the step changes increased in amplitude as the tracking task was made more difficult (ref. 10).

These studies provided valuable insights into the way in which cognitive resources are allocated in complex tasks. In addition, they established P300 amplitude as a sensitive index of the amount of processing resources, in a sense the degree of attention, that is devoted to particular classes of stimuli in complex tasks. However, possible practical applications of these results are subject to the previously discussed limitations of secondary task methodologies. Granted, the fact that measures of attention allocation can be extracted from ERPs elicited by stimuli being covertly counted, offers the possibility of applying a secondary task methodology without the need to burden the subject with additional manual response requirements (ref. 5). However, even when the stimuli being counted are embedded in the primary task, as was the case when subjects counted step changes in a cursor being tracked (ref. 10), the cognitive demands of the counting task are superfluous to the otherwise existent task demands. The question addressed by the present work was to what extent ERPs elicited by stimuli in a single, complex task, as they are processed naturalistically, will reflect the cognitive workload demands of the situation.

THE PRESENT READOUT MONITORING TASK

We designed a laboratory task which provided discrete stimuli to elicit ERPs and allowed for the manipulation of mental workload, but yet was analogous, in many ways, to the types of monitoring activities which are performed in operational environments. The richness of this task afforded the opportunity to relate the waveforms elicited by similar physical stimuli to a variety of information-processing constructs, but without requiring subjects to concentrate on more than one task at a time. Our interest was in determining the extent to which graded effects on ERP amplitude as a function of mental workload could be observed within the context of this single task. Positive results will suggest the usefulness of ERPs as indicants of certain mental processes in any setting which offers the ability to time-lock recordings to a discrete eliciting stimulus, regardless of whether or not other tasks are being performed concurrently.

The Task. The subject's task was to monitor successive CRT displays of a circular array of six two-digit readouts. On each presentation of the display, termed a trial, one of the six readouts changed from its value on the previous trial. The values of the readouts changed, either increasing or decreasing, in large (30) or small (10) steps, within the range from 00 to 99. Large step changes were less frequent than small step changes. Presentations of the array of readouts lasted 500 msec and were separated by intervals which varied randomly from 1800 to 1900 msec.

Subjects were instructed to monitor a subset of the readouts to determine which of these readouts reached 90 or above or fell to 10 or below. Readouts which met or exceeded these target values were referred to as having gone "out-of-bounds." Workload was manipulated by instructing subjects to monitor one (low workload), two (medium workload), or three (high workload) of the six readouts. After passively monitoring a "run" of twenty trials, subjects reported the positions and sequence of occurrence of targets, i.e. attended readouts that went out-of-bounds. A given subset of readouts was designated as the targets for a sequence of six successive runs. The order of these workload conditions and the arrangement of the target readouts were counterbalanced.

In the first experiment, there was an equiprobable chance that each of the

six readouts would change on a given trial. Thus the probability of a monitored readout changing was dependent on the number of readouts being monitored. In the second experiment, monitored and non-monitored readouts changed with equiprobability, regardless of the number of readouts being monitored. Other details of the stimulus generation rules are presented in references 1, 2 and 3. A typical sequence of trials is shown in Figure 1. ERP recordings were obtained from an array of scalp electrodes with conventional methodologies (also detailed in references 1, 2, and 3).

Rationale. In the present monitoring task, the way in which the stimuli varied from observation to observation was different from the method used in most studies in the literature. Typically, the sequence of stimuli in ERP studies consists of a Bernoulli series; i.e., the particular stimulus presented on each trial is independent of that presented on previous trials. Our goal in designing the present experiments was to construct a monitoring task which called into play the same cognitive processes that are invoked in real-world monitoring tasks. In operational settings, the liklihood of a particular meter reading or display state is determined by those of the recent past; drastic changes from the last reading are less likely than relatively small changes; readings which require an overt response, e.g. because they reflect a system with some parameters "out-of-bounds," are preceded by readings in the "danger" zone.

In reflecting these features, the monitoring task used here was analogous to a wide variety of real-world challenges. A pilot's in-flight interaction with engine performance and environmental system displays or a process control operator's monitoring of plant status are fairly obvious examples of such circumstances. However, in terms of the cognitive processes invoked, the present task was also analogous, in perhaps less obvious ways, to other applied tasks. For example, an air traffic control display of planes moving about an airspace also presents information which, while not entirely predictable, is nevertheless dependent on trends. Monitoring such displays as planes move towards or away from "danger zones" and, at times, enter "out-of-bounds" conditions, such as impinging on another plane's circumscribed airspace, presents many of the same mental challenges as the present laboratory task.

This monitoring task afforded the opportunity to investigate a number of cognitive influences on ERPs. Selective attention effects on ERPs could be distinguished by comparing responses to changes in a readout being monitored as opposed to changes in a readout for which there was no such task requirement. Similarly, processing which specifically reflected the occurrence of a "target" stimulus, could be distinguished by comparing the responses elicited by attended readouts that went out-of-bounds to those elicited by attended readouts that stayed or went in-bounds, or those elicited by unattended readouts which changed in any manner. In addition, we were interested in the ERP effects related to both "tonic" changes in information processing workload, imposed by the number of readouts being monitored throughout a run of trials, and the more "phasic," dynamic influences imposed by the number of attended readouts that were close to, i.e. in "danger" of going out-of-bounds.

It is interesting to consider how the pattern of effects related to these variables, aside from demonstrating the sensitivity of ERPs to these cognitive influences, can reveal specific aspects of subjects' performance in the task. For example, the extent to which the ERPs reflect the influence of attention, the differences between targets and non-targets, or effects related to number

of monitored readouts that are "in danger" might change with the level of "tonic" workload. Will the need to monitor more readouts cause a focusing of attention, and thus perhaps greater differences between responses to monitored and non-monitored readouts? Might increasing task demands cause target stimuli to be processed differently? Might the number of readouts "in danger" be more readily noticed when workload is high, because this information could be used by the subject to distinguish which of the readouts being monitored are most likely to become targets in the near future, or will this information be disregarded when workload is high, due to the fact that there are fewer central processing resources available to devote to this additional processing?

FINDINGS

There were several aspects of the averaged ERP waveforms obtained here which showed systematic variations in response to one or more of the factors of interest. These features were designated and quantified as follows: 1) the "peak positivity" (the mean amplitude over a 200 msec epoch centered about the most positive peak between 500 and 900 msec post-stimulus onset); 2) the "slow positivity" (the mean amplitude between 900 and 1050 msec post-stimulus onset); 3) the N250 (the mean amplitude between 200 and 300 msec post-stimulus onset); and 4) the N450 (the mean amplitude between 400-500 msec post-stimulus onset). Although ERP waveshapes were generally similar across subjects, there was considerable inter-subject variability in the latency of the peak positivity. These measurement epochs were selected after inspection of across-subject, grand-average waveforms and were chosen to accommodate the systematic differences in the waveforms despite this latency variability. All of the effects discussed here were statistically significant (see refs 1, 2, and 3 for details) and were consistent between the two experiments, unless otherwise noted.

Figure 2 presents across-subject grand-average waveforms from Experiment 1 obtained from the Cz electrode. The waveforms in the two rows were sorted depending on whether they were elicited by changes in readouts being monitored or by changes in non-monitored readouts. The responses to changes that took a readout out-of-bounds are superimposed on the responses to changes that took or left a readout in-bounds. The differences among responses as a function of tonic workload can be ascertained by comparing the waveforms across columns, which present the ERPs elicited when one, two or three readouts are being monitored. The waveforms elicited by target stimuli, that is, monitored readouts that moved out-of-bounds, are presented as the dashed and dot-dashed traces in this figure.

Figure 3 presents a somewhat similar breakdown, but different layout, of the comparable data from Experiment 2. Here, the ERPs elicited under the low and high workload conditions are superimposed. In the different rows are responses recorded from different electrodes, moving from back to front of the head along the mid-line for waveforms going down the page. In the different columns are the responses elicited by changes in monitored and non-monitored readouts, both changes that took the readout out-of-bounds and those which took or left the readout in-bounds. Responses to target stimuli here are shown in the right-most column.

One other view of these data proved revealing. In Figure 4 are presented difference waveforms calculated by subtracting the ERPs obtained under the low workload condition (one readout being monitored) from those obtained under the

high workload condition (three readouts being monitored). The layout of these waveforms across the other conditions corresponds to that in Figure 3.

Target Effects. As is apparent in Figures 2 and 3, there were pronounced differences between the ERPs elicited on target trials and those elicited on non-target trials. First, responses elicited by monitored readouts as they went out-of-bounds had a much larger peak positivity than either changes in the monitored readouts that did not take the readout out-of-bounds or changes of any kind in non-monitored readouts. This effect was limited to the region of the peak positivity and probably reflects a modulation of P300 amplitude that has been reported numerous times in the past (e.g., refs 11 and 12). Interestingly, this aspect of the response to targets was present to the same extent no matter what the workload.

Second, there was an additional target effect, this one related to workload, that was evident in the difference waveforms. Figure 4 shows a negative-going wave in the 400-500 msec latency region that was present only when the responses to target stimuli elicited under low workload were subtracted from the responses to target stimuli elicited under high workload. Whether this waveform component should be seen as a negativity that enters in as the result of increased workload or a positivity that enters in as workload is reduced, **cannot** be resolved. However, the present results provide strong evidence that the workload manipulation added or enhanced a new component in the waveform, rather than simply modulating a peak, or peaks, that were otherwise there. Peaks in a difference waveform that are due to either increases or decreases in amplitude, or to shifts in latency, of peaks that are evident in the raw average waveforms, should have the same scalp distributions as those raw average peaks. Instead, Figure 4 indicates that the ERP peak in the 400-500 msec region of the difference waveforms had a more posterior distribution than either of the peaks in this vicinity of the raw average waveforms seen in Figure 3. This impression was confirmed by statistically showing that the profile of amplitudes across the scalp in this time region was different for the raw average waveforms elicited under low workload than for those elicited under high workload (ref. 3). Past references to endogenous ERP negativities in this latency region (e.g., ref. 13) provide a preliminary basis for interpreting this effect as an N450 component that is enhanced as the result of increased workload.

Selective Attention Effects. As can be seen in Figures 2 and 3, there was, at least at the low workload levels, a systematic difference between the ERPs elicited by changes in monitored and non-monitored readouts. The amplitude of the peak positivity was larger in response to changes in monitored readouts as compared to changes in non-monitored readouts. This difference is best seen by comparing the responses elicited by in-bounds changes in the monitored and non-monitored readouts. Interestingly, the attention-related effect diminished with increasing workload, apparently due more to increasing peak positivities in the responses elicited by non-monitored readouts than to those elicited by monitored readouts. This same pattern of results was found in Experiment 2, when changes in monitored and non-monitored readouts occurred with equal probabilities, and in Experiment 1, when probabilities varied with the number of readouts being monitored. The differences between ERPs elicited by monitored and non-monitored readouts at low workload may be related to selective attention differences that have been interpreted as reflecting the activation of different sensory channels (refs. 14, 15); however, the polarity and timing of this effect, and its modulation by workload, is difficult to

interpret. Further investigation of this effect is needed.

Tonic Workload Effects. Of primary concern in these data was whether there were differences in the ERP as a function of the level of workload imposed by requiring subjects to monitor different numbers of readouts. Two interactions with workload have already been noted -- with increasing workload, an N450 component emerged in the responses to target stimuli and the peak positivity increased in the responses to all changes in non-monitored readouts. In addition, two main effects of the tonic workload manipulation are evident in Figures 2, 3 and 4. First, as the subject was required to monitor an increasing number of readouts, the ERPs elicited by all stimuli showed an increased slow positivity. This slow positivity was manifest in the latency region following the peak positivity (note that the waveforms in Figure 2, which were derived from Experiment 1, span a shorter epoch than the waveforms in Figures 3 and 4, which were derived from Experiment 2) and can be seen as a slow return to baseline, but with a more posterior scalp distribution than the peak positivity itself. It is likely, although not entirely clear, that this slow positivity is the Slow Wave component which has been distinguished from the P300 on the basis of both scalp distribution and relationship to experimental manipulations (e.g., ref. 16).

A second main effect of tonic workload was apparent in the difference waveforms. When responses to readout changes from the low workload condition were subtracted from the corresponding responses from the high workload condition (Figure 4), a negative-going peak appeared in the 200-300 msec latency region. This N250 occurred in the responses to both changes in monitored and non-monitored readouts, regardless of whether these changes took the readout out-of-bounds or took or left it in-bounds. As with the N450, which was only present in the responses to target stimuli, we interpreted this effect as a negative-going component which entered or was enhanced as the result of increasing workload. This interpretation was based on the fact that the scalp distribution of this wave differed from that of the corresponding activity in the raw average waveforms and the fact that processing negativities related to selective attention have been reported in this latency region of ERP waveforms (ref. 17). Statistical tests confirmed that the amplitude profile across the scalp in the 200-300 msec latency region differed between the low and high workload conditions. To our knowledge, this workload-related effect had not been reported prior to our paper (ref. 3).

It is possible that the standing requirement to monitor a given number of readouts for minutes at a time may have caused differential DC-shifts in the EEG. The transient ERPs elicited by readout changes might then have been superimposed on different baselines, and the apparent main effects of workload on post-stimulus ERP components could have resulted from a confound of, or interaction with such differential baselines. To determine whether or not such differential pre-stimulus activity could have influenced the present findings, we did the recordings for Experiment 2 in a manner which allowed us to quantify the DC level of the pre-stimulus baselines. There were no systematic differences in the pre-stimulus baselines of the ERPs elicited under different workload conditions.

Phasic Effects of the Number of Readouts in Danger. As mentioned previously, the specific value of the readout presented on a given trial was dependent on its value on the previous trials; namely, it increased or decreased by a large or small increment from its value on the previous trial. Therefore, at any

given time, only those readouts that were within a large increment of going out-of-bounds were "in danger" of becoming targets on the next presentation. Although it was not part of the subject's defined task to attend to this aspect of the situation, and no mention was made of it in the instructions, subjects could have facilitated their performance on the task by attending to this information. Therefore, we sorted the ERPs that were elicited with different numbers of readouts "in danger," to see if the waveforms showed evidence of this factor having influenced the processing of the readouts.

Figure 5 presents the data from Cz for Experiment 1 with the responses superimposed that were elicited when 0, 1 or 2 monitored readouts were "in danger." In the two rows of waveforms are presented the ERPs elicited by monitored and non-monitored readouts. In the three columns within each half of the figure are presented the data as a function of the number of readouts being monitored -- i.e., level of tonic workload. These waveforms showed an enhanced positivity in the long latency regions with increasing numbers of monitored readouts in danger. Statistical tests (see ref. 2) confirmed this effect on the peak positivity, with the slow positivity showing the same trend but not reaching statistical significance. This increased positivity was present in both the responses to monitored and non-monitored readouts and was found to the same extent at all levels of tonic workload. When the waveforms were sorted according to the number of non-monitored readouts in danger, no systematic ERP differences were found.

These data clearly suggest that subjects processed the readouts differently depending on the number of monitored readouts that were close to going out-of-bounds, even though they were not explicitly instructed to do so. It is not clear whether this differential processing should be seen as an additional, albeit self-imposed, workload demand of the task, or whether subject's chose to assume this additional processing as a means of coping with the primary task of detecting target readouts. A number of further manipulations are necessary in order to arrive at a convincing interpretation of this effect. However, the fact that this effect occurred, suggests the value of looking more closely at subjects' strategies when dealing with non-Bernoulli sequences of stimuli.

DISCUSSION

Obviously, the monitoring task that we designed provided a rich environment for eliciting cognition-related effects on scalp-recorded ERPs. To summarize, we found:

1. An N250 wave, possibly a Processing Negativity (e.g., ref. 17), that emerged with increasing workload, in the responses to all readouts
2. An N450 wave, possibly related to the N2 complex (e.g., ref. 13), that emerged with increasing workload, in responses to the target stimuli only.
3. A peak positivity, probably related to the P300 (e.g., ref. 5), which dramatically increased in amplitude when a target stimulus occurred, increased in amplitude as a function of the number of monitored readouts "in danger," and showed an interaction with tonic workload and selective attention, such that the differences between responses to monitored and non-monitored readouts which were found at low workload levels diminished with the requirement to monitor more readouts.
4. A slow positivity, possibly related to the Slow Wave (ref. 16), which increased in amplitude with workload, in the responses to all readouts.

More work is required to determine the functional significance of the waveform changes we observed and to relate them convincingly to ERP components that have been identified in other paradigms. Nevertheless, the present findings warrant several important general conclusions. Workload-related ERP effects can be derived in single task paradigms without burdening the subject with competing task demands, the effects of different cognitive variables are specific to circumscribed regions of the waveforms, and some regions of the waveforms are affected by multiple information-processing manipulations. These relationships confirm the exquisite sensitivity of scalp-recorded ERPs to the cognitive milieu in demanding tasks and suggest the possibility of eventually indexing specific cognitive processes with specific waveform components or with the activity in specific latency regions of ERPs.

It is interesting to note, however, that even prior to attaining a thorough understanding of the functional significance of specific ERP components, one can infer, from the pattern of results, a number of indications about how subjects performed the present task. Consider the fact that changes in monitored readouts that went out-of-bounds (i.e. targets) elicited a markedly different response from changes in monitored readouts that stayed in-bounds, whereas responses to changes in non-monitored readouts did not distinguish between in-bounds and out-of-bounds changes. These results suggest that subjects did indeed selectively attend to the readout positions that they were instructed to monitor. Likewise, the fact that the ERPs showed a significant effect related to the number of monitored readouts "in danger," but no effect of the number of non-monitored readouts "in danger," suggests that subjects noticed the former but not the latter. Both of these findings are consistent with the conclusion that subjects did not process the value of non-monitored readouts despite the fact that only one readout changed on a given presentation and subjects did not know whether a monitored or non-monitored readout was about to change.

On the other hand, this conclusion must be reconciled with the fact that both the workload effect on the N250 and slow positivity, and the effect of number of monitored readouts "in danger" on the peak positivity, were found in the responses to changes in both monitored and non-monitored readouts. This finding suggests that these ERP effects reflect differential processing due to the distributing of attention among the readouts being monitored, and that this processing, in essence, is related to determining which readout changed, rather than to determining the specific value of the readout that changed. Therefore, the present ERP results can be used to infer that subjects selectively attended to the readouts that they were to monitor, that they noticed the number of monitored readouts that were "in danger" of going out-of-bounds, and that workload modified some aspects of the processing of all stimuli, whether monitored or not.

Such information would be useful to know in a number of practical applications. Design issues such as configuring display formats which minimize workload, maximizing the effectiveness of warning messages, and increasing the salience of task-critical information often hinge on reliable measures of which stimuli are being attended, whether extraneous information is intrusive, whether subjects are taking advantage of useful information that is available, and which of several alternative designs entail less mental workload. The present results point towards the possibility of using ERPs to address such issues, in situations where one can not rely on, or it is difficult to acquire, subjective and behavioral measures. Moreover, in addition to playing a

confirmatory or surrogate role, ERPs may serve a diagnostic function. When overt performance has been observed to fail, one may be able to glean information from ERP effects like those obtained here in order to indicate the particular aspects of information-processing, and by inference the particular aspects of system design, that were deficient. Beyond the design arena, such ERP measures may also be helpful for monitoring the progress of training on demanding tasks or for selecting personnel who are particularly capable of functioning in various tasks.

Of course, many of the ERP effects obtained here were small and required extensive data analysis based on average waveforms. For some engineering applications, one would have the luxury of collecting as much data and analyzing it to the extent that we did here, but in other applications one would be more constrained. Nonetheless, the present results may point the way towards other manipulations or measures that would better emphasize the effects of interest. It will be interesting to see, as studies like the present ones are recast in the operational systems or simulators whose task demands have been approximated in the laboratory, to what extent the cognitive-related patterns of ERP results become more pronounced.

REFERENCES

1. Horst, R. L., Munson, R. C. & Ruchkin, D. S.: Event-Related Potential Indices of Workload in a Single Task Paradigm. Proceedings of the 28th Annual Meeting of the Human Factors Society, 1984, pp. 727-731.

2. Horst, R. L., Munson, R. C. & Ruchkin, D. S.: Processing Inferences Derived from Event-Related Potential Measures in a Monitoring Task. Proceedings of 29th Annual Meeting of the Human Factors Society, 1985, pp. 976-980.

3. Horst, R. L., Ruchkin, D. S. & Munson, R. C.: ERP Processing Negativities Related to Workload. In R. Johnson, Jr., R. Parasuraman, & J. W. Rohrbaugh (Eds.) Current Trends in Event-Related Potential Research (EEG Suppl. 40). Amsterdam: Elsevier, 1986, pp. 186-190.

4. Donchin, E., Ritter, W. & McCallum, C: Cognitive Psychophysiology: The Endogenous Components of the ERP. In E. Calloway, P. Tueting & S. Koslow (Eds.), Brain Event-Related Potentials in Man. New York: Academic Press, 1978.

5. Donchin, E., Kramer, A. F. & Wickens, C.: Applications of Brain Event-Related Potentials to Problems in Engineering Psychology. In M. G. H. Coles, E. Donchin, & S. Porges (Eds.), Psychophysiology: Systems, Processes, and Applications. New York: Guilford Press, 1986.

6. Wickens, C. D., Isreal, J. & Donchin, E.: The Event-Related Cortical Potential as an Index of Task Workload. Proceedings of the 21st Annual Meeting of the Human Factors Society, 1977.

7. Isreal, J. B., Chesney,, G. L., Wickens, C. D. & Donchin, E.: P300 and Tracking Difficulty: Evidence for Multiple Resources in Dual-Task Paradigms. Psychophysiology, vol.17, 1980, pp. 259-273.

8. Isreal, J. B., Wickens, C. D., Chesney, G. L. & Donchin, E.: The

Event-Related Brain Potential as an Index of Display-Monitoring Workload. Human Factors, vol. 22, 1980, pp. 211-224.

9. Wickens, C. D.: The Structure of Attentional Resources. In R. Nickerson (Ed.) Attention and Performance VIII. Hillsdale NJ: Erlbaum, 1980, pp. 239-258.

10. Wickens, C. D., Kramer, A., Vanasse, L. & Donchin, E.: Performance of Concurrent Tasks: A Psychophysiological Analysis of the Reciprocity of Information-Processing Resources, Science, vol. 221, 1983, pp. 1080-1082.

11. Duncan-Johnson, D. C. & Donchin, E.: On Quantifying Surprise: The Variation in Event-Related Potentials with Subjective Probability. Psychophysiology, vol. 14, 1977, pp. 456-457.

12. Johnson, R., Jr. & Donchin, E.: On How P300 Amplitude Varies with the Utility of the Eliciting Stimuli. Electroencephalography and Clinical Neurophysiology, vol. 44, 1978, 424-437.

13. Ritter, W., Simson, R. & Vaughan, H. G., Jr.: Event-Related Potential Correlates of Two Stages of Information Processing in Physical and Semantic Discrimination Tasks. Psychophysiology, vol. 20, 1983, pp. 168-179.

14. Harter, M. R. & Aine, C. J.: Brain Mechanisms of Visual Selective Attention. In R. Parasuraman & D. R. Davies (Eds.), Varieties of Attention. Orlando, Fla.: Academic Press, 1984, pp. 293-321.

15. Mangun, G. R. R. & Hillyard, S. A.: The Spatial Allocation of Visual Attention as Indexed by Event-Related Brain Potentials. Human Factors, vol. 29, 1987, 195-211.

16. Ruchkin, D. S. & Sutton, S.: Positive Slow Wave and P300: Association and Disassociation. In A.W.K. Gaillard & W. Ritter (ed.), Tutorials in ERP Research - Endogenous Components. Amsterdam: North Holland, 1983, pp. 233-250.

17. Naataanen, R.: Processing Negativity: An Evoked Potential Reflection of Selective Attention. Psychological Bulletin, vol. 92, 1982, pp. 605-640.

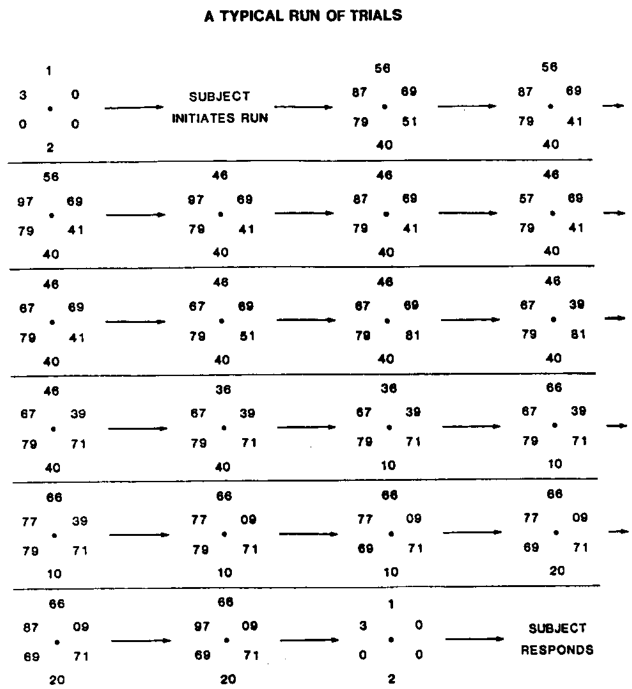


Figure 1 — A typical run of trials. Stimulus displays from a sequence of trials are shown. On each display the value of one of the readouts was different from its value on the previous display. The twenty trials are preceded and followed by a display that informed the subject as to how many and which readouts were to be monitored for "out-of-bounds" values. In this run, three readouts were monitored and the correct response at the end of the run was "3, 2, 3."

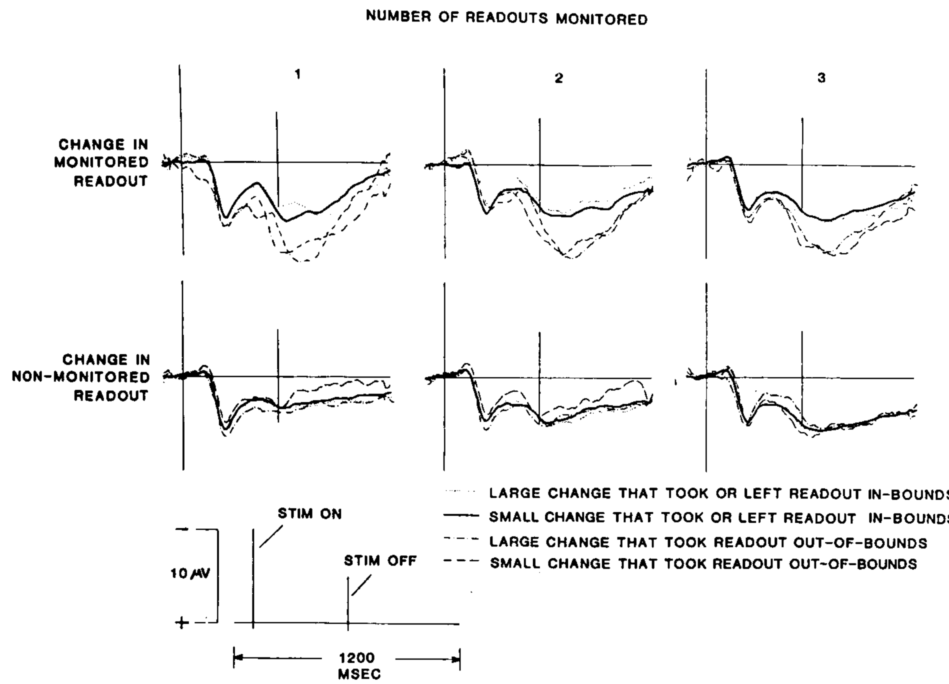


Figure 2 — Across-subject average waveforms from Experiment 1 at Cz, with responses to changes that took a readout "out-of-bounds" superimposed on responses that took or left a readout "in-bounds." Responses are sorted according to whether the eliciting change occurred in a monitored or non-monitored readout and according to the number of readouts being monitored (tonic workload).

C-3

Raw Average Waveforms

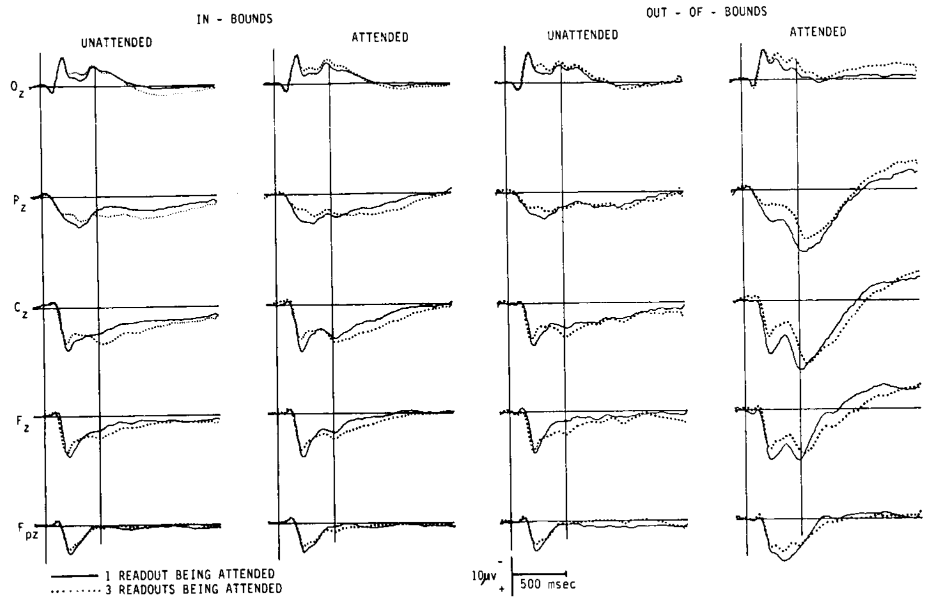


Figure 3 — Across-subject average waveforms from Experiment 2 at a range of mid-line scalp sites. Responses elicited under low workload (one readout being monitored) are superimposed on those elicited under high workload (three readouts being monitored). Responses are sorted according to whether the eliciting change occurred in a monitored or non-monitored readout and whether or not the eliciting change took the readout out-of-bounds. (Reprinted from Ref. 3)

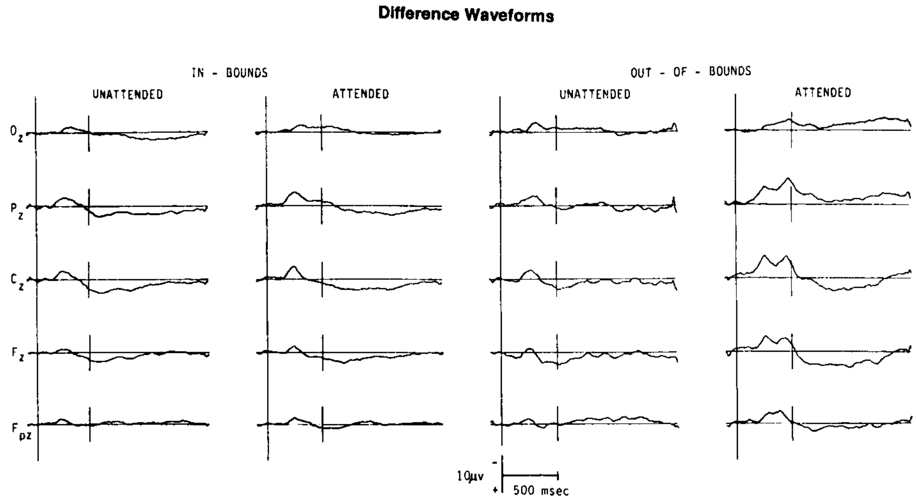


Figure 4 -- Difference waveforms corresponding to the data in Figure 3, with the responses elicited under low workload subtracted from the responses elicited under high workload. (Reprinted from Ref. 3)

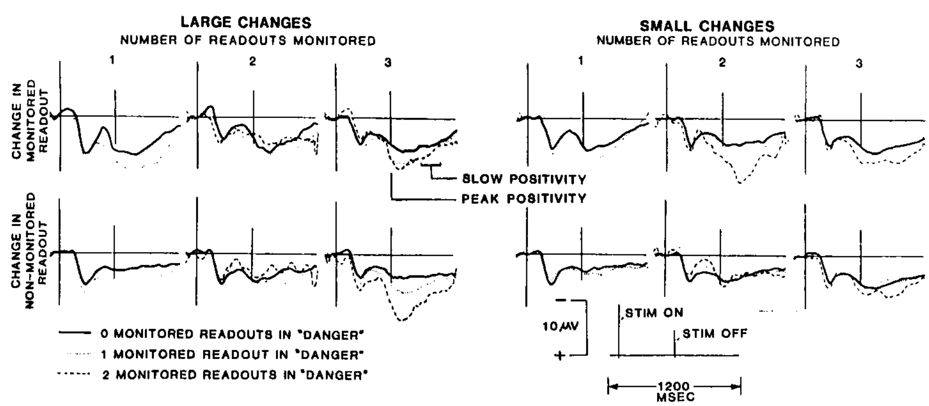


Figure 5 -- Across-subject average waveforms from Experiment 1 at Cz, with responses elicited when different numbers of monitored readouts were "in danger," i.e., within an incremental value of going "out-of-bounds." The responses are sorted according to whether the eliciting change occurred in a monitored or non-monitored readout, the number of readouts being monitored at the time, and whether the eliciting change was a large or small increment. (Reprinted from Ref. 2)

N88-23383

5/3-53
140801
18P.

BE 282249
CN

Defining and Measuring Pilot Mental Workload

Barry H. Kantowitz
Battelle Memorial Institute
Human Affairs Research Centers
Seattle, Washington 98105

Both scientists and practitioners agree that definition is a necessary precursor to productive discourse. But any definition must be clearly understood by both parties. For example, the hip musician's definition of jazz --Jazz is when you dig it, man!--does not help the naive listener who sincerely wants to appreciate jazz music but lacks the artistic sophistication of the professional musician. While this definition of jazz is too simple, the musician can also confuse a listener by excessive use of jargon that is too sophisticated. Few listeners could sympathize with a jazz trumpet player who complained about being boxed in by a C minor ninth vamp laid down by his pianist.

Similar dangers abound when research scientists try to define and explain mental workload to airplane pilots and other interested non-researchers. As a researcher I am well aware that the jargon used by human factors specialists may not always make sense to the uninitiated. Yet I also understand that an overly simple definition of mental workload --Too much mental workload is when you can't fly the plane right --also is not helpful. My goal in this article is to try to explain to the pilot why and how workload researchers approach what may appear to the pilot as a simple problem in very complex ways. There just is no easy way to define and measure mental workload.

Why Use Theory?

Researchers and practitioners can be arranged along a hypothetical continuum according to how they approach solving a problem. At the cost of only minor exaggeration we might characterize practitioners as being so anxious to solve a problem that they often solve the wrong problem whereas researchers are so anxious to get everything right that they seldom solve any problems! In order to reach a satisfactory solution, albeit not necessarily an optimal one, we must operate nearer to the middle of this continuum instead of at an extreme endpoint. It is true that an experienced problem-solver can often come up with a satisfactory answer without explicitly invoking theory. But I would argue that this approach is too idiosyncratic to work in general. The world does not have enough

experienced problem solvers to meet every need. However, one theory goes a long way. It can be applied to many different practical scenarios. Theories offer generality. We do not need a separate theory for each problem. We may not even need a very complex theory to get a direction for solving a practical problem like evaluating pilot mental workload. After all, you don't need a Ferrari to go grocery shopping. A Volkswagen will get you to the store and back. When I am asked to solve a problem like measuring pilot mental workload, I start out by looking for a handy theory. I do not expect the theory to solve my problem, only to get me started in a promising direction. Theory can be a filter that narrows down a large set of possible approaches allowing us to concentrate our efforts upon a few techniques that are most likely to yield satisfactory solutions.

There is a deplorable tendency for the practitioner to avoid theory because it does not seem relevant to the immediate problem at hand. Each problem is seen as an isolated issue and, practitioners who avoid theory run the considerable risk of reinventing the wheel time and time again without realizing it. But even the practitioner who wants to use theory must face at least two major obstacles. Most psychological theories have been formulated in arcane ways with little regard for fostering practical applications. Furthermore, there are too many theories so that it is hard for the practitioner to select one theory from the abundance created by diligent researchers. Later on I will suggest one particular kind of theory that should be useful for studying pilot mental workload. For now, I acknowledge these obstructions.

I believe that theory offers four substantial benefits to the practitioner faced with a real-world problem. First, it fills in where data are lacking. We will never have enough empirical results to solve all problems. Theory is needed for accurate and sensible interpolation. Second, theory can yield the precise predictions that engineers and designers demand. It is better to have predictions about the workload imposed on a pilot by some particular system design than to have to build the system and then obtain data to fix the next version. Third, theory prevents us from reinventing the wheel. It allows us to recognize similarities among problems. Fourth, theory is the best practical tool. Once an appropriate theory is available, it can be used cheaply and efficiently to aid system design.

Limited Capacity Theory of Attention

My approach to the practical problem of pilot mental workload is derived from basic research on attention. A detailed analysis of the kind of theory best suited for this work can be found in Kantowitz (ref.1). Here I will only summarize my conclusions in this regard. I prefer an attention theory with a single limited pool of capacity as the starting point for studies of pilot mental workload. Such a model was popularized by Broadbent (ref.2). While current views of attention realize that many of the details of this original limited-channel model are incorrect (see ref. 3 for a review), the fundamental idea of a single limited-capacity source that funds mental operations remains sound. This concept of attention is particularly useful for work on pilot mental workload because it carries with it the idea of *spare capacity.* Spare capacity is roughly defined as extra capacity not currently being used by the human but available immediately should the need arise.

There are certain assumptions used by most basic researchers studying attention and capacity that deserve explicit mention (ref. 3). First, we assume that behavior can be understood in terms of a hypothetical flow of information inside the organism. This flow cannot be directly observed but must instead be inferred from overt measures of performance. Models must not only duplicate the overt performance but must also make reasonable statements about this postulated internal information flow. For example, a female singer and a tape recording made with the proper brand of tape can both shatter a slender crystal goblet. Nevertheless, no one would claim that the human vocal tract and an electronic tape recorder produce sound by the same internal information flow.

Second, we assume that capacity is the "price" each internal processing stage charges the system to perform its own activity or information transformation. If sufficient capacity is not available, the internal processing stage may be unable to perform its function properly and/or may require greater processing time.

Third, we assume that allocation rules determine how capacity is mapped to internal stages. This is especially important when demand exceeds supply. A complete model of attention and information processing should have something explicit to say about each of these three key assumptions (ref. 3).

Defining Mental Workload

Mental workload is an intervening variable, similar to attention, that modulates or indexes the tuning between the demands of the environment and the capacity of the organism. Before considering the implications of this definition I must first explain what I mean by "intervening variable."

Intervening variables have been the subject of much discussion in psychology, especially as contrasted with hypothetical constructs (ref. 4). A hypothetical construct has surplus meaning; for example, one might try to locate the physiological basis of the hypothetical construct called the limited-capacity channel. An intervening variable is closely coupled to the operations that define it. Indeed, it ceases to exist without these operations. For example, learning is often defined as a relatively permanent change in behavior between the first test of some knowledge and a later test. Presumably better performance on the later test is evidence for the intervening variable we call learning. If the tests are removed, we can no longer make any statements about learning. Learning is thus inferred from a change in performance. It cannot be observed directly.

In a similar manner, both attention and mental workload are also intervening variables. They cannot be observed directly. We make inferences about attention or workload only on the basis of observed changes in performance. If performance decreases we often attribute this decrease to increased mental workload (or decreased attention).

There are at least four important implications of the definition of mental workload stated above. First, it implies that both underload and overload are cause for concern. In both cases there is an imbalance between the demands of the environment and the capabilities of the organism. A crew falling asleep on a trans-oceanic flight is as much a pilot mental workload problem as an engine fire. Second, the definition implies that capacity is fixed. Third, to be most useful the definition implies that spare capacity is related to mental workload and this in turn implies that a single-pool model of capacity will work better than attention models that postulate multiple sources of capacity. Fourth, it implies that the limit upon the internal information flow within the human is one of rate not amount. An analogy (ref. 5) will make this clear. No highway engineer is truly interested in the number of cars that a freeway can hold as a static

measure. While this number is important for designing parking lots, highway engineers are far more concerned with the number of cars that can flow past a given point in some specified time. Similarly, the amount of information per unit time, bits/sec, that can flow through the human is more important for understanding pilot mental workload than an absolute amount of information with no time constraint.

Measuring Mental Workload

There are three general methods for measuring pilot mental workload: (1) subjective measures, (2) objective measures, especially those based upon secondary tasks, and (3) psychophysiological measures. These are discussed in general by Kantowitz (ref. 1) and as they relate to aviation by Kantowitz and Casper (ref. 6). All methods have advantages and disadvantages. There is no clearly superior method to measure pilot mental workload in all circumstances. I believe that secondary-task measures offer the best opportunity to obtain valid and reliable indices of pilot mental workload now. In the near future psychophysiological measures may also prove to be quite useful.

The reader may be surprised that I have not endorsed subjective measures, since these are by far the most widely used method at present. While it is awfully easy to obtain subjective measures, they are quite difficult to interpret. There are at least two fundamental problems with them. First, with the possible exception of SWAT* ratings (ref. 7), the psychometric properties of most subjective rating scales have not been established. While at least interval scale properties are required for meaningful measurement and comparison, it is not at all clear that more than ordinal measurement has been achieved in most cases. Second, people are not very good at giving direct introspections that accurately reflect their own internal mental states. Psychology has long abandoned the method of introspection because it utterly failed to produce reliable data. A more recent example can be found in the work of Metcalfe (ref. 8) who studied people's ability to solve anagram puzzles and other brain teasers. Every ten seconds subjects were asked to rate on a scale of 0 to 10 how close they felt they were to a correct solution. The results were extremely lucid. People were grossly inaccurate in their ratings. When they gave high ratings, indicating that they thought they were close to a correct solution, they were more likely to give an incorrect answer than to reveal the proper solution. This demonstrates once again that subjective intuitions may not

*Subjective workload assessment technique (SWAT)

be reliable.

Thus, we are better off relying upon objective data provided by secondary tasks and psychophysiology. The secondary-task paradigm attempts to obtain direct estimates of spare capacity, and hence mental workload, by requiring an additional task to be performed at the same time as the primary flying task. Decrements in secondary-task performance are interpreted as reflecting mental workload imposed by the primary task. Primary tasks that demand greater mental workload will cause poorer performance on the concurrent secondary task.

In order for this interpretation to be valid, several control conditions must be included in the experimental evaluation of mental workload; see Kantowitz (ref. 3) for a detailed explanation and examples of published research where these safeguards have been neglected. The crucial assumption of the secondary-task method is that insertion of the secondary task does not alter primary-task performance or the internal information flow within the human operator.

In the past, secondary tasks were chosen largely on the basis of convenience with little thought given to the theoretical or methodological implications of secondary-task selection. Now, however, it is generally realized that there is no panacea that will create a universal secondary task. Many issues must be considered carefully before a satisfactory secondary task can be accomplished. Some relevant questions are:

1. Will this research be carried out in [1] an operational setting [2] a flight simulator [3] a laboratory?

2. The primary task is [1] flying [2] tracking [3] other continuous task [4] other discrete task.

3. Most primary-task information is presented [1] visually [2] auditorally [3] tactually.

4. The primary-task input information load (e.g., rate of information per unit time such as bits/sec) is [1] low [2] medium [3] high.

5. Input information load is [1] constant [2] low variability [3] high variability.

6. Output modality is mostly [1] manual [2] verbal.

7. Output responses occur [1] seldom [2] moderately often [3] frequently.

8. Operators are [1] unpracticed [2] moderately practiced [3] highly practiced professionals.

9. Operator motivation is [1] low [2] moderate [3] high.

10. Procedures associated with the primary task are [1] well-specified and usually performed in a consistent manner [2] leave the operator some discretion for arranging his work [3] vague and subject to considerable interpretation.

These considerations are sufficiently complex so that an expert system is now under construction to help choose appropriate secondary tasks. Workload COnsultant for Secondary Task Selection (W. COSTS) presents lists of questions similar to those above and makes recommendations for selecting suitable secondary tasks. This expert system uses rule-based chaining to derive its suggested secondary tasks (ref. 9).

<u>A Simulator Example of Secondary-Task Research</u>

At the risk of appearing immodest I will illustrate secondary-task techniques with a series of studies my colleagues and I have conducted in a motion-base (GAT) flight simulator at Ames Research Center (refs. 10,11,12 and 13). The primary task in all these studies was flying the simulator. The secondary task was choice-reaction time with two, three, or four alternatives. This contrasts with the typical study where a simple (one-choice) secondary reaction task has been used. However, based upon a hybrid model of attention (ref. 14) I believed that simple probe tasks were too insensitive and subject to a host of methodological problems. While many researchers felt it would be safer to use a simple probe task because this simple task would be less likely to interfere with the primary flying task, I disagreed. I believed that professional pilots would not allow the secondary task to interfere with flying. The first responsibility of a pilot is to keep the airplane safely in flight. Therefore, professional pilots seemed to me to be the ideal population for taking the risks associated with a complex choice-reaction secondary task.

Results have been excellent. Flying performance measured by root mean square error was not adversely affected by adding the complex secondary task. Furthermore, this secondary task was able to discriminate among levels of workload in many different simulated flight situations. I conclude that the choice-reaction

task should be high on everyone's list of preferred secondary tasks. Indeed, this opinion of mine is reflected in W. COSTS which tends to suggest choice reactions for almost any situation where pilot mental workload must be measured.

Psychophysiological Measures

Objective measures need not be only behavioral. The technology for recording psychophysiological correlates of behavior is now well advanced and many of these biological indicants have been used to estimate pilot mental workload (ref. 15). Once monitoring electrodes have been attached to the pilot, these indices have the advantage of being relatively unobtrusive. They do not interfere with flying as might be the case for behavioral secondary tasks. However, these data are often difficult to interpret even though they are easier to understand than most subjective ratings. Theories of psychophysiology are not yet as advanced as theories of attention and do not provide a complete framework for interpreting data.

In my laboratory we have had modest success in using heart rate (sinus arrhythmia) and evoked potential as indicants of attention in a psychological refractory period task (ref. 16) and a divided attention task described later in this volume (ref. 17). Others have successfully used psychophysiological tasks to measure pilot mental workload (see ref. 6 for a review). I believe that as theoretical models of psychophysiological indicants are refined, these techniques will become an important part of the toolbox used by human factors specialists to measure pilot mental workload.

Conclusions

The best practical tool is a good theory. Models of attention based upon a single pool of limited capacity offer an excellent starting point for measuring pilot mental workload. Thus, I define mental workload as an intervening variable similar to attention.

Objective measures are preferable for measuring pilot mental workload. Secondary tasks, especially choice-reaction time, are extremely useful in this regard. Psychophysiological tasks will be more useful in the near future as theoretical models are refined.

Acknowledgment

This research was supported by Cooperative Agreement NCC 2-228 from the National Aeronautics and Space Administration to BITS, Inc. S.G. Hart was the NASA Technical Monitor.

References

1. Kantowitz , B.H. Mental workload. In P. Hancock (Ed.), Human Factors Psychology. Amsterdam: Elsvier, 1987.

2. Broadbent, D.E. Perception and communication. Oxford: Pergamon Press, 1958.

3. Kantowitz , B.H. Stages and channels in human information processing: A limited analysis of theory and methodology. Journal of Mathematical Psychology, 29, 1985, 135-174.

4. MacCorquodale, K. & Meehl, P. On a distinction between hypothetical constructs and intervening variables. Psychological Review, 55, 1948, 95-107.

5. Broadbent, D.E. Application of information theory and decision theory to human perception and reaction. In N. Wiener & J.P. Schade (Eds.), Cybernetics of the Nervous System. Amsterdam: Elsvier, 1965.

6. Kantowitz , B.H. & Casper, P.A. Human workload in aviation. In E. Wiener & D. Nagel (Eds.), Human Factors in Aviation. San Diego: Academic Press, 1987.

7. Reid, G.B., Eggemeier, T., & Shingledecker, C.A. Subjective workload assessment techniques. In M.L. Frazier & R.B. Crombie (Eds.), Proceedings of the workshop on flight testing to identify pilot workload and pilot dynamics. Edwards Air Force Base, Ca: AFFTC-TR-82-5, 1982.

8. Metcalfe, J. Premonitions of insight predict impending error. Journal of Experimental Psychology: Learning, Memory, and Cognition, 1986, 12, 623-634.

9. Kantowitz , B.H. Workload Consultant for Secondary Task Selection. Interim report, BITS, Inc., Dec 1987.

10. Kantowitz, B.H., Hart, S.G., & Bortolussi, M.R. Measuring pilot workload in a moving-base simulator: I. Asynchronous secondary choice-reaction task. Proceedings of the Human Factors Society, 27, 319-322, 1983.

11. Kantowitz , B.H., Hart, S.G., Bortolussi, M.R., Shively, R.J. & Kantowitz , S.C. Measuring pilot workload in a moving-base simulator: II. Building levels of workload. Proceedings of the Annual Conference on Manual Control, 20, 1984, 359-371.

12. Bortolussi, M.R., Kantowitz , B.H., & Hart, S.G. Measuring pilot workload in a motion base trainer. Applied Ergonomics, 17, 1986, 278-283.

13. Kantowitz , B.H., Bortolussi, M.R. & Hart, S.G. Measuring pilot workload in a motion base simulator: III. Comparing synchronous and asynchronous secondary choice-reaction tasks. Proceedings of the 31st Annual Meeting of the Human Factors Society, 1987.

14. Kantowitz, B.H. & Knight, J.L. Testing tapping timesharing: II. Auditory secondary task. Acta Psychologica, 40, 1976, 343-362.

15. Coles, M.G.H., Donchin, E. & Porges, S.W. (Eds.) Psychophysiology. New York: Guilford Press, 1986.

16. Kantowitz, B.H. Biocybernetic analysis of a hybrid workload model. Behavioral Research Associates Technical Report BRA 85-10, May 1985.

17. Kantowitz, B.H. & Casper, P.A. Estimating the cost of mental loading in a bimodal divided-attention task: Combining reaction time, heart-rate variability and signal detection theory. NASA CP 2504, pp. 211-229, May 1988.

N88-23384

POPEYE: A PRODUCTION RULE-BASED MODEL OF MULTITASK SUPERVISORY CONTROL (POPCORN)

James T. Townsend, Helena Kadlec, and Barry H. Kantowitz
Purdue University
West Lafayette, Indiana

Abstract

Recent studies of relationships between subjective ratings of mental workload, performance, and human operator and task characteristics have indicated that these relationships are quite complex. In order to study the various relationships and place subjective mental workload within a theoretical framework, we developed a production system model for the performance component of the complex supervisory task called POPCORN. The production system model is represented by a hierarchial structure of goals and subgoals, and the information flow is controlled by a set of condition-action rules. The implementation of this production system, called POPEYE, generates computer simulated data under different task difficulty conditions which are comparable to those of human operators performing the task. This model is the performance aspect of an overall dynamic psychological model which we are developing to examine and quantify relationships between performance and psychological aspects in a complex environment.

Introduction

With increased automation in the working environment, physical demands of tasks have, in many situations, become secondary to mental or psychological demands. Automation has changed the role of the operator from one of direct control to one where the operator primarily monitors and schedules multiple tasks. This has resulted in complex systems which place greater demands on the operator's information processing capabilities. In these situations it is often assumed that performance on tasks is mediated by the allocation of processing resources which are limited (ref. 1). Mental workload is then operationally defined in relation to the overall ability of the human processing system to process information and generate responses as the task demands change (ref. 2).

Human factors and cognitive psychologists have recently begun to investigate potential variables contributing to mental workload using a variety of methods. Since mental processes are not directly observable, they are often inferred from the operator's performance or physiological measures. Alternatively, estimates of mental workload may be obtained directly from the operator's subjective judgments of the workload imposed by the task. Because of its high face validity, the latter approach of obtaining subjective ratings of workload has become widely used in human factors research.

The relationships between the performance measures and subjective ratings of workload, however, are not clear and sometimes the measures do not correlate as task demands change. In addition, many results have been accumulating (see, e.g., ref. 3 for a review), without a coherent theory to bring the observations together. Consequently, a more unified approach, which would embed the various aspects of this research area such as would be provided by a modeling approach, could clarify the relationships between performance and subjective workload measures. Our model of the complex task POPCORN, which will be described in the next section, is an attempt at this approach.

Relationships between the task type and task difficulty on the one hand, and subjective workload ratings and performance measures on the other are complex. Results seem to depend on the task itself, as well as how and when the workload manipulation is accomplished (refs. 4 and 5). Other task characteristics, e.g., task priority and reference task (ref. 6), also play a role. Most important, however, is the result of the latter study which shows that performance and workload ratings do not correlate under all conditions. Finally, while task characteristics certainly affect workload, recent investigations also seem to suggest that operator characteristics may affect not only performance but also workload ratings, at least under certain conditions (refs. 7, 8 and 9).

The considerations that are involved in examining subjective workload, some of which were briefly discussed above, underscore the importance of modeling, since from a practical, as well as a scientific view, it seems extremely important to be able to identify and quantify these various factors contributing to subjective mental workload. That is why we feel that a model, which would represent the performance as well as the psychological aspects of the operator in a dynamic way, could prove very useful in this area of research. With a working model, we could elucidate the relationships between workload (as well as other psychological) and performance measures in a quantitative way as various task characteristics are manipulated. One such possible dynamic model is shown in Fig. 1.

We began by modeling the performance component of the task. In particular, we developed a production system model of POPCORN, utilizing some of the production systems ideas developed by Newell (ref. 10) and later elaborated by John R. Anderson (refs. 11 and 12). Production systems have been useful in modeling various cognitive skills, such as general problem-solving (ref. 12) and a computer text editing task (ref. 13). Our production system will be presented following a brief description of the POPCORN task.

Description of the POPCORN Task

A complex task, called POPCORN, was recently developed at NASA by Sandra Hart for studying psychological variables that may contribute to the experience of workload. This task simulates a relatively complex automated system where the operator is responsible primarily for decision-making and the scheduling of the different components of the task in order to maximize the score in a minimum amount of time.

The POPCORN task is implemented on the IBM PC AT, and the operator interacts with it via a mouse. The complexity of a particular simulation can be manipulated primarily by the number of functions available to the operator, and ranges from level 1 (least complex) to level 5 (most complex). To begin the modeling, we chose level 2 since it has only six of the twelve functions available and thus is easier to model, yet it is psychologically interesting since some decision strategies must be employed.

The monitor display, as it appears for a level 2 scenario, is shown in Fig. 2. The larger boxes along the bottom of the display are the task boxes, with the smaller boxes beneath them used to select the different tasks. There are five task boxes, each of which will contain a task of a different type, and one penalty box which has no lid. The boxes along the right hand side are the functions used to operate on the tasks. At the second level of complexity, the functions OPEN, CLOSE, STUFF, Y->G, R->Y and SEE are available. The OPEN function opens the task box, while the CLOSE function closes it. The STUFF function is used to replace all the individual "kernels" of the task that have popped out back into their task box. The other three functions are used for kernels that have changed their state (i.e., color or visibility) in the warning zone (see below).

The scenario would proceed as follows. At specified times the task boxes are filled with the "tasks"; each task is a group of identical "kernels", the five different tasks being represented by kernels of different symbols, # - + = and *. The kernels can be released from their particular task box by first selecting that task (by moving the mouse to the smaller box underneath the task box and clicking the mouse), followed by clicking the mouse in the OPEN function box. Once the task box is open, the kernels "pop out", one at a time, and float in an upward direction at a predetermined speed specified by the experimenter. Each click of the PERFORM function (lower right hand corner of the display and available at all levels of complexity) renders one kernel of that task done, whereby the kernel disappears from the screen and the score is incremented. Only popped kernels may be performed, and only one at a time.

As the kernel moves up the screen, it may be performed as long as it has not crossed the warning line. Once the kernel

crosses the warning line, it can change its state to one of the warning states (which was predetermined by the experimenter). The "normal" state of the kernel is green. In the warning zone, it can change to either yellow, red, or invisible. As the changed kernel moves up through the warning zone, it can still be performed for points if its state is first returned to green by pressing the appropriate sequence of functions. When the kernel is returned to its green state it must first be performed before the next kernel can be operated on. These warning states are one of the ways of penalizing the operator for lagging behind. If the kernel is not performed in time, it moves to the top of the screen where it disappears and goes to the graveyard. An optional penalty for each dead kernel can be imposed by subtracting points from the score for each dead kernel.

If there is another task scheduled to enter into a task box which still has some (or all) kernels in it, the operator is given a 20 second warning by a red flashing bar under that task box. If the kernels in the task box are not done within that 20 second warning, the unperformed kernels are sent to the penalty box. There the kernels lose their identity, and since the penalty box has no lid, they begin to exit as soon as they arrive there. The points for performing these kernels are no longer obtainable; however, performing them does avoid the penalty for dead kernels.

The object of the simulation task is to obtain as many points as possible in the least amount of time. Often, therefore, the scenarios can be performed faster and more efficiently if two or more tasks are worked on simultaneously, by alternating between them. The higher levels include progressively more functions which allow the operator a wider range of options and strategies. These will not be described here since they are not included in the model at the present time. As an operator plays POPCORN, the functions and the times at which those functions are performed are stored in a response file by POPCORN.

In addition to the complexity level and also within each complexity level, the difficulty of each POPCORN scenario can be manipulated by four major variables: 1) the number of kernels in each task, 2) the total number of tasks, 3) the task schedule (i.e., the schedule of the arrival times of the tasks; a massed schedule results when all tasks arrive simultaneously, while different arrival times result in a staggered schedule), and 4) the speed of the kernels' movement. These variables will be used to examine the effects of environmental factors on the performance of POPCORN, and later to study the influences of the psychological variables of the model. We next describe the production system for the performance component of the POPCORN task.

Production System Model of POPCORN Performance

Performance of POPCORN lends itself to a production system approach since it can be readily interpreted as a hierarchy of goals and subgoals. The hierarchial goal structure is presented in Fig. 3 and the corresponding productions controlling the flow of control of the system are given in Table 1.

There are two main branches in the system. The first branch (productions P1 to P13) consists of the strategy selection that an experienced operator may engage in to prepare for playing POPCORN. Prior to the task, the operator is given a brief description of the upcoming task, called the flight plan. The flight plan provides information about the number of tasks to be done, the number of kernels in each task, the arrival schedule (massed or staggered), the speed with which the kernels move, the rewards/penalties for performed/dead kernels, and the state of the kernels in the warning zone. Based on this information and the operator's experience, (s)he can form an initial opinion about the difficulty of the upcoming scenario and decide, perhaps tentatively, on an initial strategy. The second branch (productions P14 to P44) is the production system of the actual performance of the POPCORN task. It should be emphasized that the operator is not bound in any way to use the initial strategy once (s)he starts playing. The playing strategy can be re-evaluated at any time if it is not conforming to the proper execution of the task. A demonstration of the production system follows.

The performance of the POPCORN task begins with the goal to 'play POPCORN'. Since the flight plan is the first thing to appear on the monitor, production P1 applies and the new goal becomes to 'choose an initial strategy'. If the flight plan has not yet been read and processed by the operator, production P3 applies and the goal becomes to 'read the flight plan'. Production P4 is the only one that applies here, and the operator reads the flight plan, stores the levels of the variables pertaining to the scenario (e.g., whether the speed of the kernels is slow, moderate or fast, whether the arrival schedule is massed or staggered, etc.) in working memory (WM), brings into WM the weights of these variables from long-term memory (LTM), and initializes the variable DIFF (difficulty) to zero and VARIABLE to 1. These latter two variables will be used in calculating the perceived difficulty of the scenario on which the strategy will subsequently be based.

The weights of the flight plan variables pertain to the importance of each variable in contributing to the difficulty of the scenario. For example, the speed with which the kernels move may contribute more to determining the difficulty than the number of kernels in each task, and will thus have a greater weight. Our pilot work indicates that the speed variable is the most important variable in determining the perceived difficulty of a

scenario. These weights are parameters of an operator which get updated, or tuned, based on the operator's experience. Table 2 shows a possible way of breaking down each variable into its levels, which are the independent variables of our studies by which we manipulate the difficulty or complexity of the environment. An example of the calculation of the perceived difficulty is also presented in Table 2. For illustrative purposes, the parameter values are chosen such that the DIFF variable lies between 0 and 10.

Once the flight plan is read but the strategy has not yet been chosen, production P5 applies and the goal becomes to 'weigh the variables' of the flight plan which are now stored in WM by P4. Production P7 calculates the perceived difficulty (DIFF) of the scenario in a manner analogous to the example shown in Table 2. When all the variables have been calculated into the DIFFiculty score, P8 makes the new goal to 'pick one strategy Si' (i = 1, 2, ..., 5). Here, depending on the result of the DIFF score, one of productions P9 to P13 will apply and a strategy is chosen.

The strategies are labeled S1 through S5. Strategy Si denotes that the operator will work on i tasks simultaneously. Thus, for example, when the perceived difficulty is less than 2 (i.e., a very easy scenario), production P9 will apply and the operator chooses to work on all five tasks simultaneously, strategy S5. As the difficulty increases, fewer tasks can be done simultaneously.

When the strategy is chosen, P6 and P2 return the system to the goal to 'play POPCORN' again. This time the conditions of P14 apply and the new goal becomes to 'work on the tasks'. Initially all the task boxes are closed and the kernels cannot get out. Thus if the i task boxes that the operator wants to work on are not open, P16 applies and the goal becomes 'open all i task boxes'.

Since at the second level of complexity only one task can be attended to at any one time, in this production system, task X will refer to the task the operator is currently attending to. (Note that in productions P19, P25, P26, P27 and P35 task X can also include the penalty box; however, opening or closing the penalty box constitutes an error.) From P2 task X has been tagged as the first task to be opened. But task X has not yet been selected thus P18 applies and the goal becomes to 'select task X', which is accomplished by P19 where the mouse is moved to the smaller box under task X and the mouse is clicked. When task X has been selected P20 applies and the goal becomes to 'open task X', which is accomplished by P21. When task X has been opened, but not all i tasks have yet been opened, P22 makes the next task the current task, which is then selected and opened in the same manner. Upon opening all i task boxes, P17 applies and the new goal becomes to 'work on tasks' again. Now the task

boxes are open and kernels are popping out, so P23 applies. Here the operator decides which popped kernels will have to be performed first (if $i \geq 2$). It is assumed that the task with the most popped kernels will always be chosen to be operated on first. Now the new goal becomes to 'perform popped kernels' which is where the majority of the actual playing of POPCORN takes place.

The most straightforward way to play is to select task X, if it is not already selected (P25), perform all the popped kernels of that task (P27), then select a new task with the most popped kernels (P26), perform those (P27), select another task with the most popped kernels (P26), perform those (P27), and so on until all popped kernels are done. However, other conditions may arise, particularly in faster scenarios, where the operator has to switch tasks or the order of performing the popping kernels in order to accommodate new incoming tasks without losing points or to take care of kernels that have gone into the warning zone.

If the kernels of the current task have entered the warning zone and changed to yellow, then one of productions P28, P29, or P32 applies depending on further conditions of the scenario. If there are no kernels popping out of any of the other (open) task boxes (i.e., only the current task is left to do at this point) and the scenario is not too difficult, then the operator can process the warning state and P32 applies to make the new goal to 'process the warning state'. This is the most efficient strategy in this case since a minimal amount of time is lost. Production P34 changes the top kernel in the warning zone from yellow to green, and P33 brings the system back to the goal to 'perform popped kernels' where P27 now applies. The sequence of P32, P34, P33, and P27 must be applied for each kernel in the warning zone, thus it is assumed that the warning state can only be efficiently processed in situations where there is enough time and there are no other demands on the operator. The experienced operator knows from past experience in which situations the warning states can be efficiently processed, and some pilot work has supported this assumption. For the other warning states, red or invisible, productions similar to P34 can simply be included in this part of the production system.

If the scenario is too fast (i.e., the DIFF is greater than some critical value which can be thought of as another operator parameter; here 5 is chosen somewhat arbitrarily for illustration), then P29 applies and the operator stuffs the task. This loses some time but prevents the loss of points if there is not enough time to process the warning state.

If the kernels of the current task have entered the warning zone and kernels are also popping out of other tasks, some of which may also be near or entering the warning zone, then P28 applies and the new goal becomes to 'stuff task X' in order to avoid losing them whereby their performance is postponed until

later. In this case, if the scenario is too difficult, the best strategy is to stuff the kernels back into their box and close the box (P31) in order to have sufficient time to perform the other popping kernels. If the scenario is relatively easy, then only stuffing the task (P30) may be sufficient to provide enough time to catch up with the other popping kernels.

Another situation where the straightforward sequence of selecting and performing kernels as they pop out (using productions P25, P26, and P27) may be disrupted arises when the 20 second warning flashes under a closed task box signalling the upcoming arrival of a new task in that box. In such a case, the task (called task Y in P35) has not yet been selected, and if the situation permits the processing of an additional task (e.g., when other open tasks are finished, or their kernels are popping slowly and not approaching or inside the warning zone), as judged by the operator, then production P35 applies and that task is selected and then opened (P36). In this situation the new task is incorporated into the ongoing strategy [Si becomes S(i + 1)].

If the scenario is fast and there are already many popping kernels, the operator may elect to stuff, and possibly close, one of the current tasks (P37). In this way the task box with the flashing warning in essence takes the place of one of the current tasks in the strategy, and the performance of the popping kernels can proceed in a "normal" fashion. However, the experienced operator can judge how much time is required to pop and perform the kernels, and may even be able to finish a started task before switching to the new one.

When all the kernels of the i tasks have been performed, then the best strategy is to close at least some of the finished task boxes if more tasks are expected to arrive in those boxes. If the empty task boxes remain open, then the kernels of the newly arrived task will begin to leave as soon as they arrive. Production P39 will apply in this case, and the new goal becomes to 'close (5-i) task boxes'. (Note that closing (5-i) tasks assumes that the strategy Si remains effective; this assumption seems reasonable for an experienced operator.) Again the task to have the close function performed on it must first be selected, if it is not already selected, (P43 and P19) before it can be closed (P41 and P42).

At this point the operator opens the next i tasks which contain kernels (P16) and the game continues in the same manner as above. If the operator has finished all tasks but more are to arrive later, all there is to do is wait (P38) until the new ones arrive. If all tasks are done, productions P15 and P0 end the game.

<u>POPEYE: Computer Implementation of the Production System</u>

Due to the IBM PC AT system limitations, the computer

implementation which we call POPEYE does not perform POPCORN in real time. Rather, it simulates results as if it were playing POPCORN. Each time a 'move' is executed, the scenario, as it appears to POPEYE, is updated. Thus the program keeps track of the running time, as well as the last time that the scenario was updated, and updates the scenario for the time difference. The generated responses are stored in an output file which has the same form as the replay file generated by POPCORN when a human operator is performing the task. Thus the responses generated by POPEYE can be checked by running the POPEYE output file using POPCORN's replay command.

The current version of POPEYE performs the task only under the following task constraints. 1) The schedule of task arrivals must be massed, that is, all five tasks of each set must arrive simultaneously. This was done in order to make the initial programming of POPEYE manageable. 2) The current version can only perform two sets of tasks per scenario, although it will not be a problem to make the program flexible to include any number of sets in the next version. Any warning state can be processed, and there are three different speeds available; 0.3 cm/sec, 0.7 cm/sec, and 1.2 cm/sec.

POPEYE prompts the user for a "difficulty criterion", an integer between 1 and 10. This is an operator parameter corresponding to the criterion value for the DIFF variable in the production system (which was set to 5 in Table 1 for illustration), and is used to determine if a task box should be closed after all kernels in it are done, and also to determine whether a task should be stuffed or kernels in the warning zone processed (for productions P31, and P32). This criterion is used in POPEYE by comparing it to the calculated difficulty (DIFF) of the scenario based on the flight plan variables and weights.

POPEYE also prompts the user for an operator parameter "kernels criterion", an integer between 1 and 4, which is used to determine whether to close a box after it is stuffed. If the number of kernels popped out of <u>another</u> task exceeds this criterion, the current task is stuffed <u>and</u> closed; if not, the task is only stuffed. Finally, the last prompt is for the operator's "mean to move" the mouse. This mean is used to generate an exponentially distributed random number which is added to a constant representing the minimum time between two moves.

In the current version of POPEYE the tasks are performed left to right and consecutively, unless emergency situations arise. Also all popped kernels of a selected task are completed before the next task is selected. In our pilot work, these performance assumptions were fairly well supported.

Game parameters which describe the scenario to be simulated must be provided for POPEYE. These parameters include: 1) the

number of task sets to be performed (currently only 2 are allowed); 2) the number of kernels per task (any integer between 1 and 8); 3) the schedule code; and 4) a code for the warning state. These are read from a file by POPEYE. In addition, each operator has his/her own flight plan variable weights, which are stored in a separate file. This file, in a way, represents the long term memory of the operator, and contains the weight for each flight plan variable and the weights for the different levels of each.

We have not yet analyzed the performance of the model statistically, but assessed its performance by viewing the generated results as they were replayed in the actual POPCORN task. The data simulated by POPEYE was virtually indistinguishable from data produced by human operators. Depending on the parameters given, POPEYE can generate data which result in performance that looks either like a well-practiced operator or a beginner.

Future Directions

The next version of POPEYE will aim toward a dynamic interactive model which will include such psychological variables as frustration, motivation, and working memory, as shown in Fig. 1. Throughout the report, some reference was already made to some of these psychological variables, and in fact the current version of POPEYE already contains and uses some of these variables (e.g., working memory), albeit not very formally at this stage of modeling. Thus the extension toward a dynamic psychological model is a very natural consequence of our work so far.

By studying the performance aspects of POPCORN as they change with different psychological manipulations, for example, by increasing the number of frustrating events or errors that the operator experiences, we can examine how these psychological variables contribute not only to the operator's performance but also to his/her experience of the individual aspects thought to underlie workload experience such as time pressure, physical and mental effort, etc. In addition, we can investigate how these individual aspects contribute to an overall experience of workload. In this way, POPEYE can be extremely useful in the investigation of the interactions of these (and possibly other) psychological variables with the performance component of the model and their contribution to the experience of workload.

With the exception of Madni and Lyman (ref. 14), no one to our knowledge has attempted to model mental workload and its relationships with performance and task characteristics. Madni and Lyman's model is an extended Petri net representation by which they attempt to describe and quantify task-imposed workload. However, we are not aware of a computer implementation of their petri net model. Petri nets are similar to production

systems in that they are formal models of information flow. Whereas both approaches rely on some matching of conditions to proceed from one state to another, production systems additionally postulate a hierarchial structure of goals which governs the overall behavior. The goal structure seems to be more appropriate to model the goal-directed behavior of human operators.

Thus, the production system approach is a useful and suitable representation of POPCORN performance. It is straightforward, and simply by adding more productions it can be fairly easily expanded to model higher levels of complexity. Also, since an action of an operator at any given time only depends on the current state that he/she finds him/herself in -- that is, the transition from one state to another depends only on the current state and not on any of the previous states -- the production system can be naturally generalized to a state probabilistic model by employing a Markov process approach.

The dynamic model will also be very useful in estimating workload ratings under different environmental conditions. For example, a straightforward estimate of workload may be obtained by simply estimating the absolute number of productions required to complete the task. Alternatively, a more complex and accurate estimate may result from a weighted combination of the productions, where a production with more conditions to be matched or more consequents to be performed may contribute to a greater extent. In summary, we feel that this approach to the modeling of POPCORN and employing the model to predict workload ratings is very useful and holds much promise.

Acknowledgment

This research was supported by Workload and Performance Research Group, NASA-AMES, grant # NAG 2-307, to the first author.

REFERENCES

1. Navon, D., & Gopher, D.: On the economy of the human-processing system. Psychological Review, 86, 1979, pp. 214-255.

2. Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? Human Factors, 26, 1984, pp. 519-532.

3. Moray, N.: Subjective mental workload. Human Factors, 24, 1982, pp. 25-40.

4. Yeh, Y.-Y., Wickens, C. D., & Hart, S. G.: The effect of varying task difficulty on subjective workload. Proceedings of the Human Factors Society, 29th annual meeting, 1985, pp. 765-769.

5. Thornton, D. C.: An investigation of the "Von Restorff" phenomenon. Proceedings of the Human Factors Society, 29th annual meeting, 1985, pp. 760-764.

6. Gopher, D., Chillag, N., & Arzi, N.: The psychophysics of workload - a second look at the relationship between subjective measures and performance. Proceedings of the Human Factors Society, 29th annual meeting, 1985, pp. 640-644.

7. Damos, D.: The relation between the Type A behavior pattern, pacing, and subjective workload under single- and dual-task conditions. Human Factors, 27, 1985, pp. 675-680.

8. Damos, D., & Bloem, K.: Type A behavior pattern, multiple-task performance, and subjective estimation of workload. Bulletin of the Psychonomic Society, 23, 1985, pp. 53-56.

9. Salvendy, B. (Ed.): Responses of Type A and Type B individuals performing a supervisory control simulation. Proceedings of the Second International Conference on Human-Computer Interaction, pp. 67-74. The Netherlands: Elsevier Science Publishers, 1986.

10. Newell, A.: Production systems: Models of control structures. In W. G. Chase (Ed.), Visual information processing. New York: Academic Press, 1973.

11. Anderson, J. R.: Language, Memory and Thought. Hillsdale, N.J.: Erlbaum, 1976.

12. Anderson, J. R.: Acquisition of cognitive skill. Psychological Review, 89, 1982, pp. 369-406.

13. Card, K. S., Moran, T. P., & Newell, A.: Computer text-editing: An information-processing analysis of a routine cognitive skill. Cognitive Psychology, 12, 1980, pp. 32-74.

14. Madni, A. M., & Lyman, J.: Model-based estimation and prediction of task-imposed mental workload. Proceedings of the Human Factors Society, 27th annual meeting, 1983, pp. 314-317.

Table 1

Production System for Performing POPCORN

```
P0:  If the goal is to play POPCORN
          and all tasks are finished,
     then pop the goal and END !!!

P1:  If the goal is to play POPCORN
          and the flight plan is presented and not read
          and an initial strategy has not been chosen,
     then the subgoal is to choose the initial strategy.

P2:  If the goal is to choose an initial strategy
          and the strategy has been chosen,
     then tag task X as the first task to begin working on
          and press 'return' on the keyboard,
          and pop the goal.

P3:  If the goal is to choose an initial strategy
          and the flight plan has not been read,
     then the subgoal is to read the flight plan.

P4:  If the goal is to read the flight plan,
     then read the flight plan
          and store the levels of the individual variables
             LEVEL(VARIABLE) in working memory (WM)
          and bring in the weights of the variables
             WEIGHT(VARIABLE) from long-term memory (LTM) to WM
          and initialize DIFF = 0, VARIABLE = 1
          and pop the goal.

P5:  If the goal is to choose an initial strategy
          and the flight plan is read and processed,
     then the subgoal is to weigh the variables.

P6:  If the goal is to weigh the variables
          and the strategy is tagged as chosen,
     then pop the goal.

P7:  If the goal is to weigh the variables
          and VARIABLE < 6,
     then DIFF = DIFF + LEVEL(VARIABLE) * WEIGHT(VARIABLE)
          and VARIABLE = VARIABLE + 1.

P8:  If the goal is to weigh the variables
          and VARIABLE ≥ 6,
     then the subgoal is to pick one strategy Si.
```

Table 1 (con't.)

P9: If the goal is to pick one strategy Si
and $0 \leq$ DIFF < 2,
then put strategy Si = S5 in WM
and tag the strategy as chosen
and pop the goal.

P10: If the goal is to pick one strategy Si
and $2 \leq$ DIFF < 4,
then put strategy Si = S4 in WM
and tag the strategy as chosen
and pop the goal.

P11: If the goal is to pick one strategy Si
and $4 \leq$ DIFF < 6,
then put strategy Si = S3 in WM
and tag the strategy as chosen
and pop the goal.

P12: If the goal is to pick one strategy Si
and $6 \leq$ DIFF < 8,
then put strategy Si = S2 in WM
and tag the strategy as chosen
and pop the goal.

P13: If the goal is to pick one strategy Si
and $8 \leq$ DIFF ≤ 10,
then put strategy Si = S1 in WM
and tag the strategy as chosen
and pop the goal.

P14: If the goal is to play POPCORN
and the strategy is chosen
and tasks are available for play,
then the subgoal is to work on the tasks.

P15: If the goal is to work on the tasks
and no tasks are available for play
and no more tasks are expected to arrive,
then pop the goal.

P16: If the goal is to work on the tasks
and the strategy is to work on (i) tasks simultaneously
and (i) tasks with kernels have not been opened,
then the subgoal is to open (i) task boxes.

P17: If the goal is to open (i) task boxes
and (i) boxes are open,
then pop the goal.

Table 1 (con't.)

```
P18: If the goal is to open (i) task boxes
          and less than (i) boxes have been opened
          and task X is not selected,
     then the subgoal is to select task X.

P19: If the goal is to select task X,
     then move the mouse to task = X
          and click the mouse
          and pop the goal.

P20: If the goal is to open (i) task boxes
          and task X is selected
          and task X is not open,
     then the subgoal is to open task X.

P21: If the goal is to open task X,
     then move the mouse to function = OPEN
          and click the mouse
          and pop the goal.

P22: If the goal is to open (i) task boxes
          and less than (i) boxes have been opened
          and task X is open,
     then tag task X as the next new task (i.e., X = new task).

P23: If the goal is to work on the tasks
          and (i) task boxes are opened
          and kernels are popping out,
     then tag task X = task with the most popped kernels
          and the subgoal is to perform popped kernels.

P24: If the goal is to perform popped kernels
          and all kernels from the open task boxes are finished,
     then pop the goal.

P25: If the goal is to perform popped kernels
          and task X is not selected,
     then the subgoal is to select task X.

P26: If the goal is to perform popped kernels
          and task X is selected
          and task X has no popped kernels
          and task X' is open and has popped kernels,
     then tag X = X'
          and the subgoal is to select task X.
```

Table 1 (con't.)

```
P27: If the goal is to perform popped kernels
            and task X is selected
            and task X has popped kernels
            and the top kernel is green,
     then move the mouse to function = PERFORM
            and click the mouse.

P28: If the goal is to perform popped kernels
            and kernel(s) of task X is (are) in the warning zone
            and other kinds of kernels are also popping,
     then the subgoal is to stuff task X.

P29: If the goal is to perform popped kernels
            and kernels of task X are in the warning zone
            and no other kinds of kernels are popping
            and DIFF > 5,
     then the subgoal is to stuff task X.

P30: If the goal is to stuff task X
            and DIFF < 5,
     then move the mouse to function = STUFF
            and click the mouse
            and pop the goal.

P31: If the goal is to stuff task X
            and DIFF > 5,
     then move the mouse to function = STUFF
            and click the mouse
            and move the mouse to function = CLOSE
            and click the mouse
            and pop the goal.

P32: If the goal is to perform popped kernels
            and the kernels are in the warning zone
            and no other kinds of kernels are popping
            and DIFF < 5,
     then the subgoal is to process the warning state.

P33: If the goal is to process the warning state
            and the top kernel is green (i.e., warning state is
                processed),
            then pop the goal.

P34: If the goal is to process the warning state
            and the top kernel is yellow,
     then move the mouse to function = Y->G
            and click the mouse
            and pop the goal.
```

Table 1 (con't.)

```
P35: If the goal is to perform popped kernels
          and a 20 sec. warning is flashing under closed task Y
          and other popping kernels are not in or near the
               warning zone
          (and task Y is not selected),
     then tag task X = Y (Y = task with warning flashing)
          and the subgoal is to select task X.

P36: If the goal is to perform popped kernels
          and a 20 sec warning is flashing under task X
          and task X is selected
          and task X is not open,
     then the subgoal is to open task X.

P37: If the goal is to perform popped kernels
          and a 20 sec warning is flashing under task Y
          and kernels of task X are popping "too fast",
     then the subgoal is to stuff task X.

P38: If the goal is to work on the tasks
          and no tasks are available for play
          and more tasks are expected to arrive,
     then wait for the new tasks.

P39: If the goal is to work on the tasks
          and (i) task boxes are opened
          and all kernels of these (i) tasks are finished
          and more tasks are expected to arrive into those boxes,
     then the subgoal is to close (5-i) task boxes.

P40: If the goal is to close (5-i) task boxes
          and (5-i) task boxes are closed,
     then pop the goal.

P41: If the goal is to close (5-i) task boxes
          and (5-i) task boxes are not closed
          and task box X is open (and empty) and selected,
     then the subgoal is to close task X.

P42: If the goal is to close task X,
     then move the mouse to function = CLOSE
          and click the mouse
          and pop the goal.

P43: If the goal is to close (5-i) task boxes
          and (5-i) task boxes are not closed
          and task X is not selected,
     then the subgoal is to select task X.
```

Table 1 (con't.)

P44: If the goal is to close (5-i) task boxes
and (5-i) task boxes are not closed
and task X is closed,
then tag task X = new task to close.

Table 2

The Variables, Weights, and Levels of the Flight Plan

Variable	Description	Weight		Levels	Weight(level)	
1	# of tasks to do	θ_1	(2)*	5 tasks	a_1	(0.25)*
				10 "	a_2	(0.50)
				20 "	a_3	(1.0)
2	# kernels/task	θ_2	(1)	2 kernels	b_1	(0.6)
				4 "	b_2	(0.8)
				8 "	b_3	(1.0)
3	speed of kernels	θ_3	(5)	slow	c_1	(0.1)
				moderate	c_2	(0.5)
				fast	c_3	(1.0)
4	arrival schedule	θ_4	(1)	massed	d_1	(0.8)
				staggered	d_2	(1.0)
5	warning state	θ_5	(1)	none	e_1	(0.0)
				yellow	e_2	(0.5)
				red	e_3	(0.75)
				invisible	e_4	(1.0)

* Note: The numbers in brackets are example values used in the example calculation below.

Example: Suppose the scenario to be played contains 10 tasks each with 4 kernels/task; the speed is moderate, the arrival schedule is staggered, and the kernels turn yellow in the warning zone. Then

$$\begin{aligned} \text{DIFF} &= \theta_1 a_2 + \theta_2 b_2 + \theta_3 c_2 + \theta_4 d_2 + \theta_5 e_2 \\ &= 2(0.5) + 1(0.8) + 5(0.5) + 1(1.0) + 1(0.5) \\ &= 5.8 \end{aligned}$$

=====> choose strategy S3.

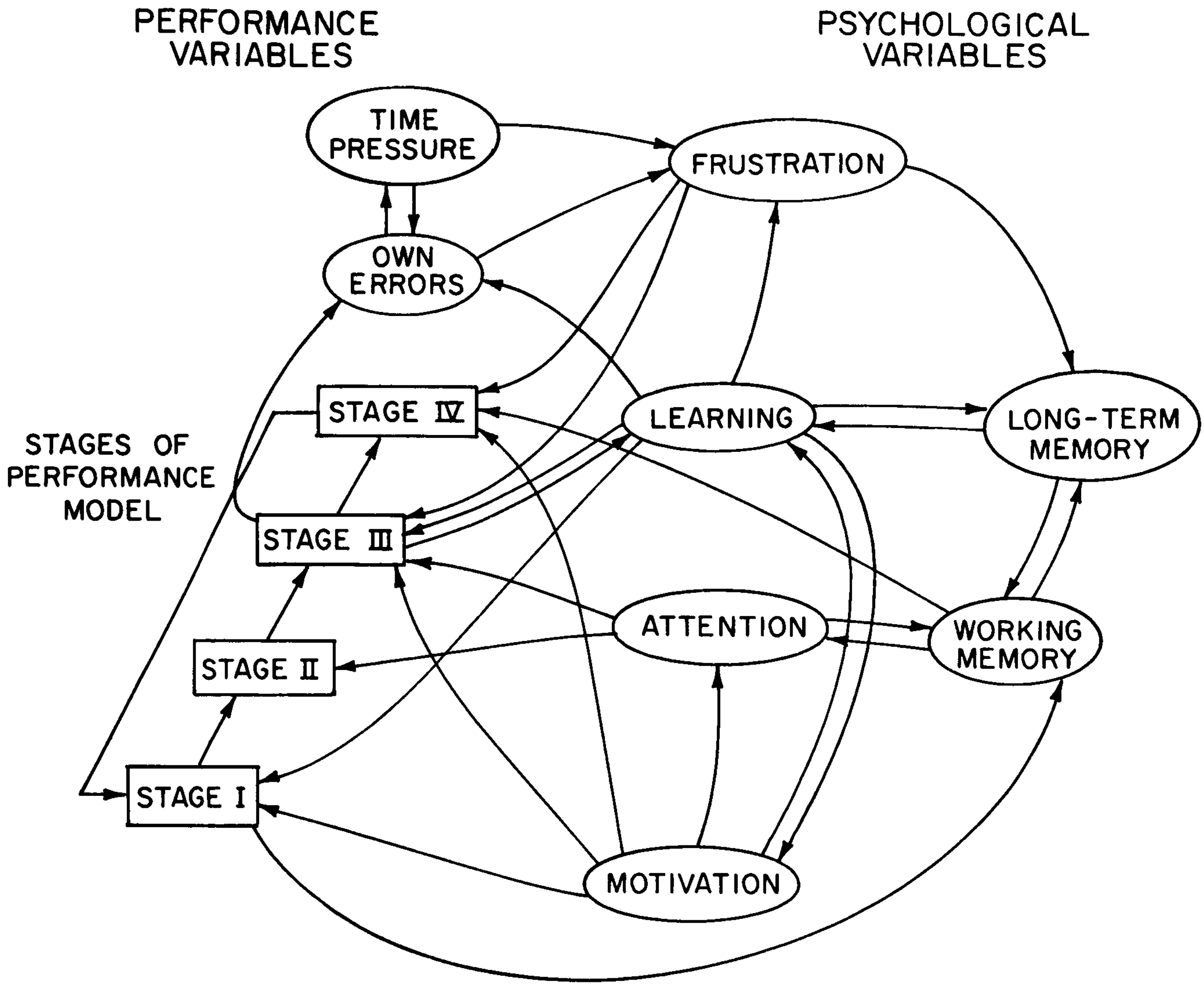
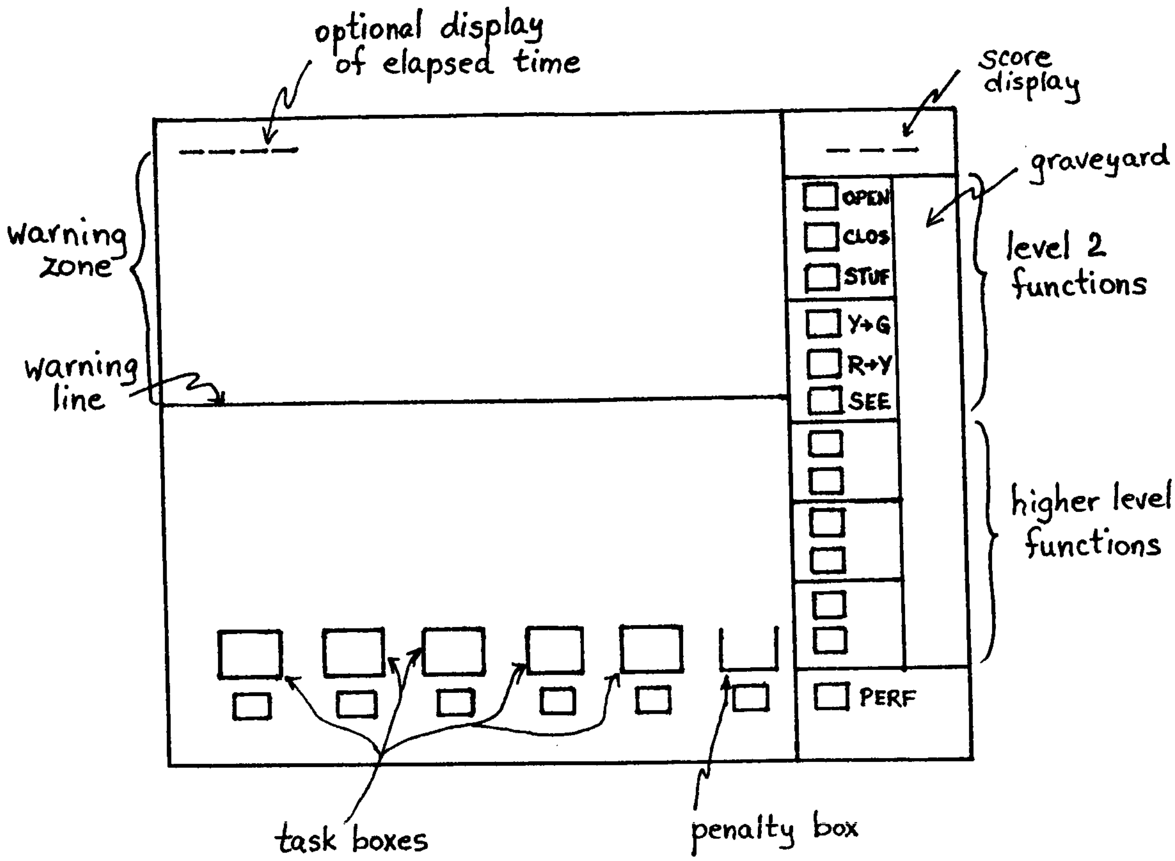


Figure 1. A dynamic psychological model showing the possible reciprocal relationships between the performance component of the model and the psychological variables.



<u>Figure 2.</u> Monitor display of the POPCORN task at the second level of complexity.

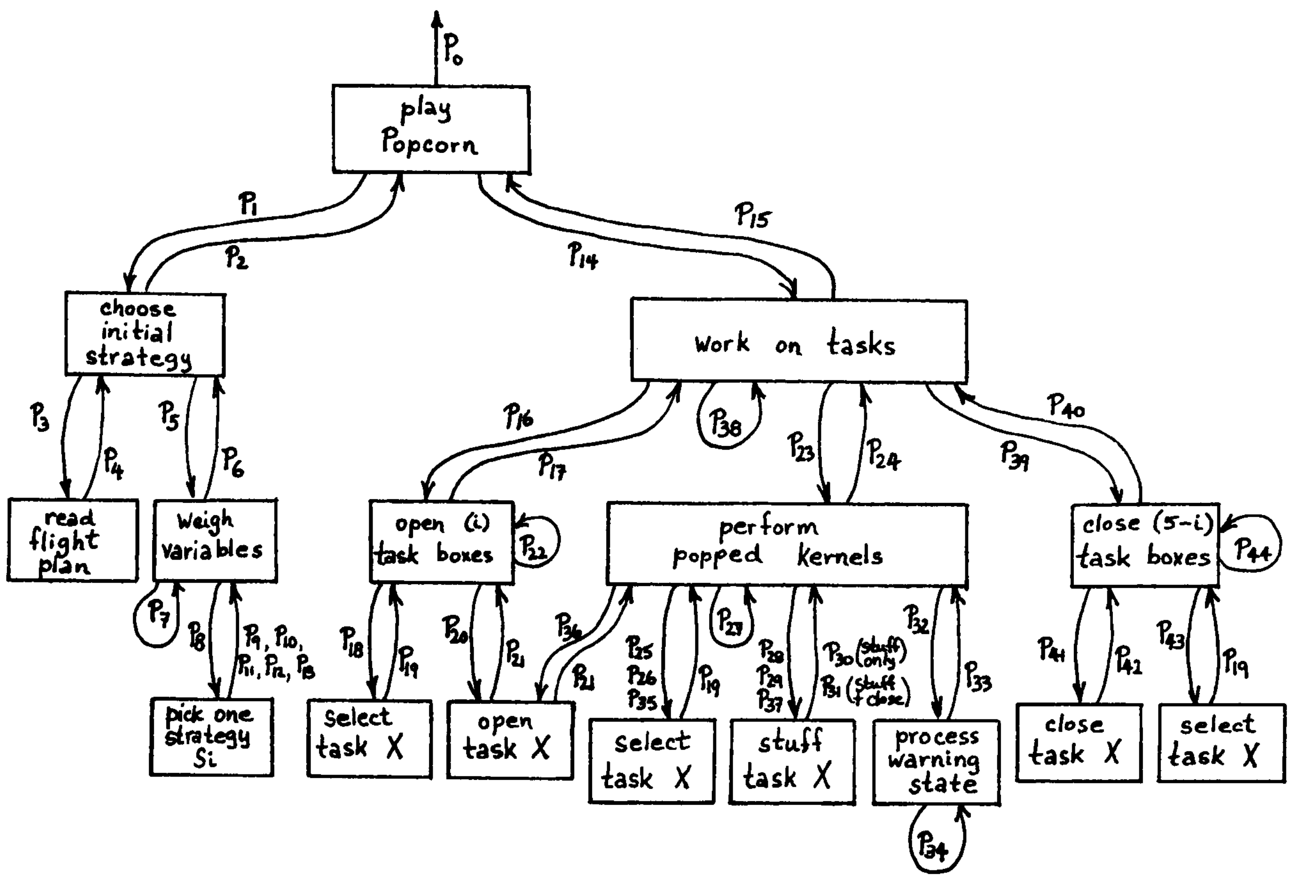


Figure 3. The goal structure for the production system of the performance component of POPCORN.

N88-23385 5/5-53 140 X03 19 P.

ESTIMATING THE COST OF MENTAL LOADING IN A BIMODAL DIVIDED-ATTENTION TASK: COMBINING REACTION TIME, HEART-RATE VARIABILITY AND SIGNAL-DETECTION THEORY

Patricia A. Casper and Barry H. Kantowitz
Purdue University
West Lafayette, IN

P 9 / 9109 - CN

The topic of workload has drawn considerable interest in the field of ergonomics for a number of years. For as long as man has been at work researchers have been concerned with quantifying the amount of load, or physical stress, placed on him. Advances in automation and technology however have recently changed the nature of man's work from that of physical laborer to mental laborer, shifting the primary focus from the human's physical capabilities to the level of cognitive or mental load with which the human can effectively cope. Estimation of a worker's ability to handle a mental task has revealed itself to be a more complex undertaking than the analogy originally suggested.

Many techniques have been used, some successfully and some not as successfully, in the effort to determine the nature and extent of the cost to the human operator for performing cognitive work. In general, methods can be classified into three broad categories, most of which will be addressed in this paper. The categories are: performance measures, subjective measures, and physiological measures. Performance measures assume that the operator's interactions with the system will result in different levels of performance depending on the difficulty of the task. Thus, such measures reflect whether or not the operator is able to meet the demands of the task. Increased task difficulty will manifest itself in the form of increased errors and slower reaction times. Unless secondary task methodology is used, however, these measures do not provide any indication of how much **spare** capacity the operator may have to perform additional tasks.

Subjective measures are based on the assumption that an operator is able to evaluate his own level of workload and thus these measures utilize a set of questionnaires on which the operator rates his degree of load. In addition to being convenient, subjective techniques are diagnostic, and often reveal sources of workload attributable to an operator's internal characteristics such as motivation, frustration, etc.

Physiological measures are based on the premise that mental tasks are performed at a certain physiological **cost** to the operator, with indications of load showing up in a number of observable physiological systems. The list of indicators is long, and includes measures of heart rate, heart rate variability, respiratory activity, blood pressure, body temperature, galvanic skin response, direction of eye movements, urochemical analysis, pupil diameter, muscle tension, and

event-related cortical activity (ERP's). The most obvious advantages of the physiological measures over the rest are their relative objectivity, their ability to be recorded continuously, and their unobtrusivity in operational settings. Since the greater portion of workload research being done today is directed at the operator at work (pilots, in particular), the unobtrusivity of these measures stands out as one of their most attractive features. Of popular interest are the measures of cardiac functioning, which will be the focus of this paper.

Mental workload has been shown to be a multidimensional construct reflecting the interaction of many factors, including an operator's training and skill level, task demands, as well as the operator's physiological state, which itself is a function of manifold homeostatic systems. To prove reliable, an approach to mental workload estimation must be malleable to the dynamic nature of the concept of workload itself.

As an example, suppose I wished to evaluate the level of frustration of a subject performing a difficult versus an easy war-type video game. Further, suppose that I employed two different dependent variables - number of enemy "hits" and heart rate. When the results of the "experiment" are analyzed I find that the difficult game produces a much higher heart rate in the subject than does the easy game, but the number of hits is the same for the two. This point illustrates the fact that different measurement devices are sensitive to different components of workload - physiological measures tap operator strain or effort (not to mention physical load), and performance measures reflect on the difficulty of the task. It may very well be that the two games were both too easy or both too hard, revealed by the fact that performance was the same on both. Nonetheless, the performance measure has told me nothing of the subject's level of frustration during the two tasks.

In the search for measures useful both in the laboratory and in operational environments it is highly unlikely that one approach, or measuring stick, will provide all the answers, since what is being measured is a dynamic and multifaceted concept. Careful definitions of mental workload paired with careful selection and implementation of a number of metrics are currently the most promising of steps toward a solution. Since the rigors of defining mental workload have been covered elsewhere in this volume, this effort will focus on a review of several approaches to the study of mental load using cardiac measures, and on the combination and interpretation of several metrics from different classes in a divided attention task performed in the laboratory.

Relationship of physiological systems to cognitive systems

According to Hancock (ref. 1), "If ERP's represent the highest scoring physiological measure on the scale of spatial and systemic congruence with respect to CNS activity, then measures pertaining to heart rate and its derivatives are currently the most practical method

of assessing imposed mental workload".

Before beginning a review of studies employing cardiac measures of load there are several important issues that need be addressed. The first of these major questions facing the scientist using physiological measures of cognitive processing concerns the exact relationship between the physiological systems and the cognitive systems. The term system is used here to represent a highly complex inter-connected network of processes that are constantly changing and approaching a goal that is oftentimes unknown. How do the physiological systems respond to different levels of cognitive processing? Is there really a physiological cost to thinking? Although perhaps more obvious to those using physiological measures, the relationship problem is nonetheless present in every approach to quantifying workload.

A widely held biological conception is that the physiological processes are in constant oscillation seeking a homeostatic state that will balance input from environmental factors, self-generated information, task-specific information, and biological functioning (refs. 2 and 3). The forecast for someone trying to measure the physiological cost associated with varying levels of cognitive load is grim from this perspective, since the physiological systems are "programmed" towards homeostasis and will adjust what parameters are necessary to keep things in even keel. It is possible that overall system output could remain the same due to the operator not performing a required task, or by the adoption of strategies altering the level of performance of several tasks. The physiological system keeps itself in a state of preparedness for emergencies by storing a certain level of "reserve capacity" to be used only in extreme cases (ref. 3). Situations most likely to allow use of the reserve capacity include extremely fearful or stressful situations, extreme physical loads, extremes of temperature, etc. These are not the situations normally encountered in a laboratory experiment; therefore, few studies should show physiological correlates of mental load. A quick glance through the literature will show that this is not the case. Many studies report changes in physiological processes associated with manipulated changes in mental load. Unfortunately, the problem is quite the opposite - the influence of **too many** variables is evident in cardiac records. One technique, however, the spectral decomposition of the heart inter-beat interval into its constituent frequency components, shows the most promise for looking at, if not unconfounding, the variances associated with a number of different physiological systems. This promising avenue will be explored later in this report.

Factors associated with cardiac output

Once one is willing to accept the idea that physiological processes are an accurate reflection of implicit mental processing, one must also realize that cardiac functions are also affected by a number of factors not thus far known to be related to cognition. Documented correlates include age, temperature, emotions, physical

load, level of responsibility, level of task-related risk, respiration, and noise (refs. 4 and 5). Even in the most carefully conducted laboratory experiment many of these factors are difficult, if not impossible to control. The state of affairs worsens as one considers the current interest in applying measures of workload in operational environments where even less control is possible.

Grain of analysis

As with other measures of workload, an issue of debate is the unit of measurement, or grain of analysis used in recording and summarizing data. Research has shown that different results may be found depending on whether data (reaction time, d') are averaged over all of the trials within a block or conditional upon the types of trials comprising a block (only one response required, two responses required) (refs. 6 and *). The three measures to be discussed in this paper differ in the amount of data that is collapsed over, with mean heart rate spanning the most, followed by overall heart rate variability, followed lastly by spectral analysis. A number of researchers have expressed concern over studies reporting data based on summary statistics for heart rate data inherently based on a non-random time series (refs. 7 and 8).

Related to the grain of analysis problem is the issue of whether cardiac responses to levels of tasks or to components of tasks should be observed (ref. 4). Should data be averaged over a block of trials of the same task (e.g. difficult mental arithmetic vs easy mental arithmetic) or over similar parts of a task occurring across trials (e.g. stimulus perception, mental rotation, etc.)? Clearly, those interested in operator responses to overall levels of mental load (that is, ergonomists) are interested in the first question. Any indicator sensitive to varying levels of task load is useful to someone with that purpose in mind. But to the cognitive psychologist, who is interested in discovering the architecture of the processing system, the second alternative appears more attractive. Ultimately, all researchers, basic and applied, are interested in a priori prediction of workload levels given certain task combinations. Thus, the major problem has two parts. A detailed analysis of laboratory tasks used in workload studies must be first undertaken, so that the components comprising a given task may be clearly specified. This would be followed by examination of cardiac responses associated with each component (e.g. perceptual input, central processing, and response processing) of the task. Only then can predictions be made concerning workload levels inherent in untested combinations of the examined task components.

The next sections will present a critical review of several studies using each of the cardiac measures of workload - mean heart rate, overall heart rate variability, and spectral analysis of heart rate.

---

*Casper, P.A. (1986) A signal detection analysis of bimodal attention: Support for response interference. Unpublished Master's Thesis. Purdue University.

Mean heart rate

Unless stated otherwise, it is assumed that HR is measured offline. Although there are some recent developments in online measurement techniques,* most research reports data that were collected as interbeat interval scores and subsequently analyzed offline, although ECG's provide a visual report of the data during the experiment (ref. 9).

As mentioned previously, mean HR makes the least parsimonious use of the available heart inter-beat interval data of the three measures. The overall statistic of HR is computed as 1/IBI (in seconds). Most studies using mean HR as a dependent variable take an average of the HR over each task period or experimental condition. Some studies, however, report second-by-second levels of mean HR (collapsed across trials and subjects) so that an approximation of the complete waveform may be seen. Such an approach is to be preferred to condition means since it is known that HR is extremely variable during the first few seconds of a task and may contaminate the data from the rest of the recording interval. Plots of the overall trend can be observed and outlying data removed from subsequent analysis.

Lacey's intake-rejection hypothesis

The majority of experiments reviewed were directed at supporting or providing evidence against Lacey's intake-rejection hypothesis (ref. 10). Specifically, Lacey proposes that an acceleration in HR accompanies tasks requiring complex "internal" processing such as mental arithmetic or memory scanning. Accordingly, HR deceleration accompanies tasks requiring attention or responses to external stimuli. The cardiovascular system is presumed to exert an influence on the bulbar-inhibitory area of the brain, which serves to enhance or inhibit detection of sensory inputs. Such responses are said to be biologically adaptive in that a faster HR is effective in shutting out potentially distracting noise so that the internal processing may proceed unhindered. HR deceleration supposedly reduces internal noise, enhancing signal detection sensitivity. Such a process would result in faster reaction times and increased accuracy to stimuli.

In the earliest of the reviewed studies addressing the intake-rejection hypothesis, Kahneman, Turskey, Shapiro, & Crider (ref. 11) observed mean HR, pupil diameter, and skin resistance to phases of a task in which subjects added 0, 1, or 3 to each of 4 serially presented digits, and reported the transformed series. Although task difficulty effects were seen only in the skin resistance and pupillary measures, all measures reflected an increase in the phase of the task where the digits were mentally manipulated, followed by a peak and sharp decline in the response phase, supporting Lacey's hypothesis. Problematic for the experiment is a trend towards differences in the

---

* Adie, P., & Drasic, C. (1986) Validation of a mental workload measurement device. Unpublished master's thesis. Department of Industrial Engineering, University of Toronto.

dependent variables among the three levels of difficulty conditions prior to any procedural differences in the tasks (i.e. prior to digit presentation).

In a more common manipulation of attentional direction, Coles (ref. 12) instructed subjects to search a 40 x 60 letter array for targets either highly discriminable or not easily discriminable from the background letters. The targets were the letter "e" or the letter "b", distributed with varying density among the letter "a" distractors. Detected targets were either counted (internally-directed attention) or denoted by a check mark (externally-directed attention). Support for Lacey's hypothesis was found, since decreased target letter discriminability resulted in decreased HR (and increased HR deceleration), and counting targets caused HR to decelerate while checking targets caused HR to accelerate. As with the Kahneman et al. (ref. 11) experiment, pre-search task differences in mean HR for the two search conditions overshadowed the findings, not to mention the fact that physical workload was also greater in the externally-directed attention condition where the subjects checked each target detected. Also, complete testing of Lacey's hypothesis was not possible due to the unavailability of reaction time data (except in the form of # of lines searched) in the task. As mentioned previously, decreased HR producing enhanced sensitivity for externally-presented stimuli should be reflected in reaction time and accuracy in the task. No error data were reported in the study.

The major argument for an alternative explanation of cardiac acceleratory and deceleratory changes involves the level of verbalization involved in the tasks (ref. 13). Presumably, "intake" tasks are associated with a higher level of internal verbalization than are "rejection" tasks. Klinger, Gregoire, & Barta (ref. 14) measured mean HR, rapid eye movements (REM's), and electroencephalogram alpha levels (EEG) in tasks where subjects performed mental arithmetic, counted aloud by two's, indicated preferences between two activities, mentally searched among alternatives, imagined a liked person, or suppressed thoughts of a liked person. The levels of HR found in the study were, from highest to lowest, in the order of the tasks just given. Tasks associated with the three highest levels of HR involved both concentration (internal processing, or rejection tasks, according to Lacey) and verbalization. Thus there appears to be a plausible (and more parsimonious, according to some) explanation for the observed set of data.

Elliott (ref. 13) has criticized Lacey's intake-rejection hypothesis and studies supporting it. Besides claiming that there is a general lack of empirical support for the hypothesis, (a disputable claim, upon surveying the literature) he further argues that the hypothesis is untestable due to the lace of sufficient operational definitions. A more parsimonious account, he suggests, is Obrist's conception of a cardiac-somatic relationship (ref. 15), where HR changes are attributed to motor activity. In this sense, HR is used as a response, and not as a cause of changes in processing efficiency. This leads the discussion to the arousal model, to be reviewed next.

Arousal models versus mental load models

The Yerkes-Dodson Law predicts an inverted U-shaped function relating performance on a mental task to the level of arousal, or stress befalling the performer. Zwaga (ref. 16) argues that the concept of arousal is a better account of observed HR changes during an experiment. Zwaga gave his subjects a paced mental arithmetic task consisting of five minutes of rest, six minutes of the arithmetic task, and five more minutes of rest. Heart rate during the first minute of the task was the highest, and thus was discarded. He further found that HR during the task was higher than that during the rest periods, but that HR decreased with the duration of the task period. HR also declined with each session of the experiment, even when the sessions were separated by a 24 hour period. Although a mental load model would predict higher HR during the task period than in rest, such a model has no explanation for why HR continued to decrease throughout the task period and with further sessions. Such findings are easily accommodated by an arousal model that predicts eventual habituation to repeated presentations of stimuli.

Cacioppo & Sandman (ref. 17) maintain that the level of cognitive demands of a task, and not a general level of sympathetic arousal, are the reason underlying observed HR effects. In their experiment, subjects were given either problems to solve (anagrams, arithmetic, or digit-string memorization), or slides of autopsies to look at. The autopsy slides were associated with two levels of stressfulness, with low stress slides being pictures taken from a distance of an accident victim, and high stress slides being close-ups of badly-mutilated accident victims. The assumption was made that stressfulness was equivalent to unpleasantness, with difficult cognitive tasks being rated as more unpleasant or stressful than easy cognitive tasks. Measuring only the first five heart beats in each task condition, difficult (stressful) cognitive tasks were associated with higher HR than easy cognitive tasks, while the stressfulness of the autopsy slides did not affect HR. Averaging over difficulty, cognitive tasks produced an increase in HR, while autopsy slide viewing produced a decrease. An arousal hypothesis would have predicted increased generalized sympathetic responses to the stressful autopsy slides relative to the low stress slides, and increased overall HR to the autopsy slides relative to the cognitive tasks. Since this was not found the authors concluded that mental processing demands associated with cognitive tasks are responsible for observed HR changes. The conflict between the two competing hypotheses could possibly be resolved by equating the measurement procedures (discarding obviously outlying HR scores obtained in the first few minutes of a session).

Laboratory versus field findings

Two of the reviewed experiments observed HR in operational environments, and found virtually no changes associated with mental load. This finding is surprising compared with the wealth of evidence supporting the use of HR to measure mental load in the laboratory. Melton, Smith, McKenzie, Wicks, & Saldivar (ref. 18) studied mean HR, urine steroid, epinephrine, and norepinephrine levels, and level of anxiety in air traffic control (ATC) workers employed at low traffic

control centers. In contrast to findings of studies at high-density traffic centers, no HR increases from off duty to on duty were observed in the ATC workers.

A comprehensive study evaluating 20 different workload measures, including HR and heart rate variability (HRV), was conducted by Wierwille & Connor (ref. 19) using a simulator in three levels of flight difficulty. Of the physiological measures studied, only mean pulse rate was observed to increase monotonically with imposed flight difficulty. No effects on HRV (scored by the standard deviation) were observed. Subjective measures, followed by performance measures, were the most sensitive to imposed load.

Hart & Hauser (ref. 20) found that the level of pilot responsibility (left seat versus right seat) and the segment of flight were able to produce changes in mean HR. HR was higher for the pilot in control of the plane than for the co-pilot, and was higher during take-off and landing phases segments compared to segments of level flight. A major problem with field studies, even if observed changes in HR are observed, is the lack of environmental control. A useful distinction among types of stress has been suggested, and that is the consideration of informational versus emotional stress. Presumably an operational environment, especially in flight, would contain more levels of emotional stress than that encountered in a laboratory, while informational stress could potentially be the same in the two environments. An experiment by Sekiguchi, Handa, Gotoh, Kurihara, Nagasawa, & Kuroda (ref. 21) in which six tasks were used ranging from tracking in the laboratory to an actual flight task supported such a notion. Perhaps the arousal hypotheses, although not useful in the laboratory environment, holds potential for testing in operational environments.

<u>Heart rate variability</u>

The major problems facing researchers using heart rate variability, or sinus arrhythmia, as a dependent measure are associated with 1) the choice of a valid and sensitive scoring method, and 2) how to remove (or prevent) contamination of observed results by influences unrelated to cognitive processing, e.g. physical load, respiration, etc.

Data scoring

Statistics used to estimate the degree of variability among a collection of IBI scores include the typical standard deviation, the number of reversals (points of inflection) in the HR signal (ref. 22), the frequency that the HR signal crosses the mean or 3, 6, or 9 beats per minute on either side of the mean (ref. 23), and the mean square of successive positive or negative (or both) differences (MSSD) between the heart rate signal. Essentially, the various scoring methods differ as to how much data are collapsed over, and whether amplitude or frequency information is included in the calculation. A comprehensive review of factor and spectral analytic techniques is provided by Opmeer (ref. 24).

Since so many empirical factors are allowed to vary, even when the selection of a scoring method is held constant, no particular statistic emerges as best in any given situation. There is some indication, as will be discussed in the section on spectral analysis, that those methods accounting for the direction and amplitude of change in the IBI are the most sensitive.

Physical versus mental load

It has been typically observed that increases in imposed physical load elevate mean HR while increases in imposed mental load decrease HRV. Such effects have often been obscured, however, due to the employment of a binary choice task at differing rates of stimulus presentation as a manipulation of task difficulty. Such a treatment confounds levels of mental load with levels of physical load. Unfortunately in some cases this confound can "cancel out" HRV effects actually due to increased mental load. Kalsbeek & Sykes (ref. 25) used such a procedure and failed to find HRV differences between levels of task difficulty.

In a classic study, Boyce (ref. 26) factorially manipulated levels of physical and mental load in an attempt to separate effects on HRV (measured by the standard deviation) associated with the two factors. Subjects were given a one- versus two-digit mental arithmetic task in which they had to move a pointer (attached via a cable to a weight) to the correct answer. Physical load was varied by changing the heaviness of the weight attached to the end of the cable. Results indicated an increase in mean HR due to both physical and mental load, while HRV decreased with increases in mental load and increased with increases in physical load.

Inomata (ref. 27) found no HR or HRV differences among rest periods and periods of a visual search task characterized by four levels of memory load, and no differences between those measures among the four load conditions. HRV was scored using the standard deviation and the sum of the frequencies per minute crossing the mean or 3, 6, or 9 beats per minute away from the mean. When the data were re-analyzed after removing data associated with overt body movement (subject's moving in their chairs, etc.), only the second deviation score decreased with increasing memory load.

Using a more complex statistic, Luczak (ref. 28) gave subjects a binary choice reaction time task with and without physical load. HRV was scored by dividing all of the positive differences (in rate) between successive heart beats by the frequency of relative maxima and minima in the time series. Physical load was achieved by having subjects move various parts of their body at the same time as they performed the binary choice task. They found that HR was correlated highly with motor load, while HRV was correlated with mental load. HRV decreased with increasing task difficulty.

Despite a confound with physical load, Ettema & Zielhuis (ref. 23) found increased HR, blood pressure, and respiration and decreased HRV with increasing levels of mental load achieved using a paced binary choice task at 20, 30, 40, and 50 signals per minute. The

heart rate, blood pressure, and respiration measures were all positively correlated with each other, and negatively correlated with both measures of HRV. HRV was scored as either the frequency of HR above or below 3, 6, or 9 beats away from the mean, or as the sum of the absolute differences between successive levels of HR.

Spectral analysis of heart rate variability

Unlike the two methods just discussed, which focus on the overall variability of the cardiac signal, the spectral analysis technique treats the IBI data as a time series upon which analysis methods in the frequency domain or the time domain may be applied. Debate has arisen concerning the appropriateness of using the typical analysis of variance statistics, which assume random samples, on non-random data. Specifically, Luczak & Laurig (ref. 8) have pointed out that when such statistics are used on time series data of IBI's the degrees of freedom associated with the experimental conditions are overestimated. This is because the samples are not random and reflect the interaction of many rhythmically occurring functions in the autonomic nervous system. It is obvious to most that the overall mean or variance of such a series does not reflect the rhythmicity of the underlying processes. Two alternative procedures remain: analysis methods from the time domain, and analysis methods from the frequency domain.

Time domain methods

Methods in this class involve the shifting of a time series in time by a specified amount of lag, and then either correlating the signal with itself (autocorrelation) or with another series (cross-correlation), in order to see power trends in the data. Since there is a great deal of noise present in the series, noise that is usually not of empirical interest, it must be removed before the factors of interest can be examined. Noise removal techniques are complex and are discussed in further detail in Coles et al. (ref. 29). In general, time domain methods have been left to scientists in electrical engineering, with psychologists choosing to employ more traditional analysis techniques.

Frequency domain methods

Analysis of heart rate variability in the frequency domain shows the greatest promise among all the cardiac measures as a reliable indicator of operator workload. Despite its methodological and theoretical promise, fewer papers have been published using this method than the two previously discussed, no doubt due to its greater complexity. These techniques, known as spectral analysis, or harmonic analysis, break the cardiac signal down into its constituent frequency components. Conceptually this is similar to the way total variance is partitioned into that accounted for by main effects and interactions in an analysis of variance (ref. 9). First, the series is transformed into one sampled at equal intervals (since most data are a measure of the R-R interval, which varies), and then a Fourier

analysis is performed which reveals the amplitude of the variance at each frequency of the signal. The sum of the energies in each interval is equal to the overall variance of the IBI. Partitioning the variance, or energy, in this way allows the researcher to see the effects of a manipulation on the individual components of the cardiac signal, even if those effects can't be controlled for in the first place. Although it is considered a more elegant technique than the others, use of the technique alone is no substitute for careful experimental design to minimize influences from sources other than those of interest. Experiments should be designed to minimize potential confounds from rhythmically-occurring biological processes that are not specifically related to cognitive processing per se, such as the time of day, ambient temperature, etc.

Different biological functions contribute power to different frequencies of the total cardiac output. The results from experiments using spectral analysis of IBI data usually reflect a body temperature component at about 0.05 Hz, a blood pressure component around 0.1 Hz, and a respiratory component in the area between 0.25 and 0.40 Hz, the normal adult breathing rate of 15 - 24 breaths per minute (ref. 30). In addition, a component may appear around the same frequency as the task presentation rate. If the task were a binary choice task with stimuli presented once every 2 seconds, a task-related component might occur at 0.5 Hz. Such a phenomenon has been called "entrainment", and refers to the synchronization of certain internal rhythms with external ones. The effect arises due to HR deceleration just prior to an expected stimulus, and acceleration just after stimulus presentation. There is also evidence that blood pressure can be entrained by respiration if the respiration rate is high and deep (ref. 31).

Not all researchers have shown the same degree of concern for the influences of respiration on the distribution of power in the cardiac spectrum. Mulder & Mulder (ref. 30) intentionally manipulated subjects' frequency and depth of respiration alone and while engaged in cognitive tasks. Results indicated that frequency bands toward the low end of the spectrum (e.g. 0.06-0.14 Hz) were not at all affected by respiration, while moving up the spectrum found effects of both frequency and depth. Increasing the difficulty of cognitive tasks was found to decrease the power inherent in a frequency band around 0.1 Hz relative to other frequency bands. Mulder & Mulder described the power at 0.1 Hz as an indicator of the amount of time spent in "controlled processing".

Spectral techniques have also been used in environments other than the laboratory. One study used tasks ranging from bedrest to treadmill exercise to tracking and actual flight that showed the power in the 0.1 Hz range to increase with moderate mental load, and decrease with increases in mental load (ref. 32). In the flight task, power in the .1 Hz range increased in the preflight check and decreased during takeoff and landing, a result complemented by HR studies (ref. 20).

One operational environment in particular, however, has turned up

results contrary to those found in flight environments. Egelund (ref. 33) reports that most studies of driving find that HR decreases with the number of hours driven, while HRV tends to increase, presumably due to fatigue. The physical work associated with maneuvering a vehicle in traffic contributes to increases in HR. Nygaard and Schiotz (ref. 34) had subjects drive a 340 kilometer course on either straight flat highways or ones with many hills and turns. They found no difference in HRV (as measured by single deviant heartbeats) between the two types of roads. Suspecting insensitivity of their measure, among other factors, Egelund (ref. 33) reanalyzed Nygaard and Schiotz's data using spectral analysis of the interbeat interval data, HRV (the standard deviation), and mean HR. Egelund predicted that the 0.1 Hz region of the spectrum would reflect an increase over the amount of time driven, while HR would decrease over time. No changes in HR or HRV were found as a function of distance driven, however, a slightly significant increase in the variability in the 0.1 Hz region was found for 2 of the last 5 segments of the journey. Although the results supported those from an earlier study, their statistical weakness was blamed on a number of factors, namely, the shortness of the test drive, and driver experience. It is worthy to note that 4 of the 8 subjects had had their licenses for two and one-half years or less (one had even had hers for only 2 weeks).

Earlier in this paper some of the problems associated with using the usual summary statistics on time series data were mentioned. A possible solution to this problem has materialized in the form of a summary statistic appropriate for spectral analytic techniques, called the weighted coherence (ref. 9). The statistic is useful for correlating the power variations at one frequency with those at another. This would allow the power variability at the respiratory frequency to be correlated to the variability at the 0.1 Hz frequency, for example. Currently it is possible to do a cross-spectral analysis, where the coherence (similar to r^2) of one rhythm with another at one specific frequency can be determined. However, without prior knowledge of which exact frequencies are of interest it was not possible to get this statistic to apply to a range of frequencies. The proposed measure, the weighted coherence, is an indication of the total variance shared by two rhythms within a limited frequency band. Finally, a means of summarizing across frequencies is available, although Porges and his colleagues did not report data validating the statistic.

The divided attention experiment

Next we will report on an experiment carried out in our laboratory combining performance and physiological measures of workload. Since the data were only recently collected, the findings reported are preliminary and much work remains to be done.

The task employed was a bimodal divided attention task in which subjects simultaneously attended to two streams of discrete stimuli, and responded manually to changes in one modality and vocally to changes in the other modality. The events in the auditory modality

were high or low-frequency tones lasting 100 msec, with 1100 msec allowed for response after tone presentation. The visual events were 100 msec flashes of a red or green light, with the same response interval as for the auditory task. A sequence of events lasted for 160 trials, or about 3.2 minutes. Subjects were instructed to respond as quickly as possible via either a keypress or by saying the word "diff" into a microphone, each time they observed a signal in a modality that was different from the previous signal in that modality. Half of the subjects used a vocal response to the auditory channel and a manual response to the visual channel, while for the other half of the subjects the response requirements were reversed. It should be noted that the response mappings for the former group should lead to better performance, since input and output modalities are more compatible for the auditory task than those used by the latter group (ref. 35). Tasks employing multiple modalities are useful in that they parallel tasks in operational environments more than the more traditional laboratory tasks, both in their difficulty and in their multimodal nature.

Task difficulty was manipulated by varying the number of tasks simultaneously performed (one = single stimulation, two = double stimulation), and the degree of synchrony between two tasks. In the synchronous case, the auditory and visual stimuli occurred simultaneously, with a total of 1100 msec allowed for the subject to respond to both of the tasks. In the asynchronous case, presentation of the auditory or the visual sequence was delayed by 300 msec after that in the other modality. Presumably, tasks that occur asynchronously in each modality are easier to perform since attention may be switched between the two and responses need not necessarily be executed simultaneously.

Dependent variables were reaction time (RT), d' and beta (response criterion), and heart rate. For the first three measures, the data were examined both on an overall basis, and conditional upon the type of trial in the other modality: no response, response. Several cardiac measures were calculated, including mean HR, HR variance, mean successive differences in HR, variance of successive differences in HR, and the variability in the .1 Hz region of the power spectrum.

Performance measures

Not surprisingly, RT reliably distinguished between the easy and difficult levels of the task, with scores being fastest during single stimulation, and slowest during double stimulation. There is no a priori reason to suspect a difference in RT's between the auditory lagged and the visual lagged conditions, and there was none found. In general, as has been previously found, RT's to the visual channel were faster than those to the auditory channel. The visual RT advantage was most evident during the easier (one task lagged) versions of the task than during the more difficult task where auditory and visual stimuli were presented simultaneously. Subjects responded more quickly with practice, and were faster when the response modalities were compatibly arranged than when incompatibly arranged.

D' scores were not significantly different in the easy and difficult versions of the task, although the trend was in the right direction, with d' slightly higher in the easy condition. Contrary to the RT results, d' was higher for the auditory than for the visual channel, however the pattern was the same as the RT results with the auditory d' advantage being greater during the asynchronous tasks than during the synchronous task. A compatible response modality for the auditory channel also produced higher d' scores than the incompatible arrangement. Given conflicting RT and d' results we intend to examine the reaction time density functions to see if the response for one modality was always executed before that to another modality, or if sometimes the response order traded off between the two modalities. Such data should reveal whether capacity was shared between the two (dependent processes) or reallocated to the other task once a task was completed (independent processes).

Values of beta were lowest in the synchronous condition, and more comparable between the two asynchronous conditions. Beta was also highest for whichever modality used a vocal response. This measure is useful in distinguishing increased performance from merely a lowered subjective criterion to respond, as opposed to a true increased sensitivity to the signal events. As was expected, the most difficult condition, the synchronous condition, resulted in the lowest values of d' (although not significant), paired with the lowest values of beta, indicating that even though the criterion to respond was lowered the subjects could still not effectively distinguish the signals from the noise.

Previous experiments in this series have shown there is an asymmetric trade-off of performance between the auditory and the visual channels dependent on whether or not 1) a response is made in the other channel, and 2) whether or not that response is overt (hit) or implicit (correct rejection) (ref. 7). Performance in the auditory channel is best when there is no overt response made to the visual channel, and worst when there is an overt response to the visual channel. Performance in the visual channel has not been shown to be affected by events in the auditory channel, for reasons beyond the scope of this paper. Further breakdowns of the data show that the visual response events causing the auditory performance decrement are both hits and false alarms, implicating interference between the channels at the response stages of processing.

At the present time we are able to report data for RT conditioned on whether or not there was a response in the opposite channel. RT was significantly faster when no response (either a hit or a false alarm) was executed in the opposite channel. The interaction of trial type with modality revealed that the RT advantage on no response trials was shown only for the visual channel. The frequency differences between the high and low tones are suspect for causing this apparent departure from earlier findings.

Cardiac measures

At the time of this report, HR data was available for 6 of the 24 subjects run in the experiment. Mean HR scores showed a decrease in HR throughout the experiment. Of HR, HR variance, mean successive difference in IBI's (MSD), and variance of successive difference in IBI's, only mean HR reflected differences between the pre-task baseline period (82 BPM) and the task period (76 BPM). HR did not distinguish, however, between the single and double stimulation versions of the task.

HR variance was significantly greater during the last half of the experiment than in the first half, but decreased within a half, perhaps reflecting the fact that subjects were growing increasingly fatigued and exerting greater effort during the portions of the experiment between rest periods.

Although not significant, the MSD measure was positive (reflecting decelerating HR) during the baseline period and negative (reflecting accelerating HR) during the task period. MSD variance did not show any effects of any of the experimental manipulations.

The IBI data were subject to interpolation to create a regularly-sampled sequence, and were input to a spectral analysis program revealing the density at each frequency in the spectrum. The power in four different frequency bands was examined: 0.06-0.14 Hz, 0.16-0.24 Hz, 0.26-0.32 Hz, and 0.34-0.42 Hz (ref. 30). Analysis of variance did not reveal differential sensitivity of the four frequency bands to manipulations of task difficulty. Several factors may account for the null findings. Although it seems plausible that our divided attention task should be at least as difficult as those reported previously using HRV as a measure, it is possible that it was not so difficult as to cause differing degrees of **effort** in the subjects. No performance criteria were imposed on the subjects, resulting in a higher than average number of missed responses and false alarms. The signal detection measures rely on the assumption that humans are less-than perfect observers, so performance errors were not discouraged. Another possibility relates to the way the analyses were performed. Power within a band was averaged over several frequencies, possibly cancelling out any effects. Mulder (ref. 36) reported data separated into discrete frequencies that showed that the 0.06 and 0.08 frequencies in particular were the most sensitive to task difficulty. Further breakdowns of the data should either support or rule out such an interpretation, which will have to be regarded as speculation until then. Not to be excluded from consideration is the fact that 3/4 of the heart rate data has not yet been analyzed, implicating insufficient power in the present null results.

Future experiments will also examine phasic HR, in a manner similar to the experiments reported earlier by Kahneman et al. (ref. 11) and Coles (ref. 12). The divided attention task has potential as a task using longer trials such that cardiac responses during different segments of a trial may be observed.

General conclusions

The importance of addressing mental workload as a multi-dimensional construct cannot be overemphasized. The potential for interactions among metrics used to assess load and the degree of imposed load is great and oftentimes unpredictable. The importance of two factors is evident: careful experimental design, and a grain of data analysis appropriate to the characteristics of the monitored signal.

Separating overall variability into smaller parcels allows us to observe the interrelationships among the different biological systems as they are related to mental processing. For physiological systems at least, the closer the data resemble continuous data, the better. At this point it seems clear that even though apparently extraneous influences can be observed and documented, they cannot be removed. Since a human is a complex system, complex responses to external and internal demands will be reflected in empirical data. Spectral analytic techniques are extremely powerful and useful tools for assessing external attentional demands placed on operators, but use of them will not guarantee solution of the workload evaluation problem. No matter what degree of experimental control is exercised over an experiment, the operator at work is going to be under a number of uncontrolled, and perhaps even unknown influences, all of which interact dynamically to result in a given level of operator strain. Nonetheless, fractionization of the task components, as well as the associated measures of workload and performance, appears to be the surest path to the study of understanding the nature of the interaction.

ACKNOWLEDGEMENTS

This research was supported by Cooperative Agreement NCC 2-228 from the National Aeronautics and Space Administration, Ames Research Center; S. G. Hart was the Technical Monitor.

REFERENCES

1. Hancock, P. A. (1985) Physiological reflections of mental workload. Aviation, Space, and Environmental Medicine, 56, 1110-1114.

2. Cannon, W. B. (1932) The wisdom of the body. New York : W. W. Norton.

3. Kalsbeek, J. W. H. (1973) Do you believe in sinus arrhythmia? Ergonomics, 16(1), 99-104.

4. Firth, P. A. (1973) Psychological factors influencing the relationship between cardiac arrhythmia and mental load. Ergonomics, 16(1), 5-16.

5. Hart, S. G. (1986) Theory and measurement of human workload. In J. Zeidner (Ed.), Human Productivity Enhancement: Training & Human Factors in System Design, vol. 1, pp. 396-456. New York : Praeger.

6. Casper, P.A. & Kantowitz, B. H. (1985) Seeing tones and hearing rectangles: Attending to simultaneous auditory and visual events. Trends in Ergonomics/Human Factors II, in R. E. Eberts & C. G. Eberts (Eds.), Elsevier Science Publishers.

7. Luczak, H., & Laurig, W. (1973) An analysis of heart rate variability. Ergonomics, 16(1), 85-97.

8. Porges, S. W., Bohrer, R. E., Cheung, M. N., Drasgow, F., McCabe, P. M., & Keren, G. (1980) New time series statistic for detecting rhythmic co-occurrence in the frequency domain : The weighted coherence and its application to psychophysiological research. Psychological Bulletin, 88(3), 580-587.

9. deBoer, R.W., Karemaker, J.M., & Strackee, J. (1985) Description of heart-rate variability data in accordance with a physiological model for the genesis of heartbeats. Psychophysiology, 22(2), 147-155.

10. Lacey, J. I. (1967) Somatic response patterning and stress: Some revisions of activation theory. In M. A. Appley & R. Trumbull (Eds.), Psychological stress: Issues in research, New York: Appleton-Century-Crofts.

11. Kahneman, D., Turskey, B., Shapiro, D, & Crider, A. (1969) Pupillary, heart rate, and skin resistance changes during a mental task. Journal of Experimental Psychology, 79(1), 164-167.

12. Coles, M. G. H. (1972) Cardiac and respiratory activity during visual search. Journal of Experimental Psychology, 96(2), 371-379.

13. Elliott, R. (1972) The significance of heart rate for behavior: A critique of Lacey's hypothesis. Journal of Personality and Social Psychology, 22(3), 398-409.

14. Klinger, E., Gregoire, C., & Barta, S. (1973) Physiological correlates of mental activity : Eye movements, alpha, and heart rate during imagining, suppression, concentration, search and choice. Psychophysiology, 10(5), 471-477.

15. Obrist, P. A., Webb, R. A., Sutterer, J. R., & Howard, J. L. The cardiac-somatic relationship : Some reformulations. Psychophysiology, 6, 569-587.

16. Zwaga, H. H. G. (1973) Psychophysiological reactions to mental tasks : Effort or stress? Ergonomics, 16(1), 61-67.

17. Cacioppo, J. T., & Sandman, C. A. (1978) Physiological differentiation of sensory and cognitive tasks as a function of warning, processing demands, and reported unpleasantness. Biological Psychology, 6(3), 181-192.

18. Melton, C. E., Smith, R. C., McKenzie, J. M., Wicks, S. M., & Saldivar, J. T. (1978) Stress in air traffic personnel : Low-density towers and flight service stations. Aviation, Space, and Environmental Medicine, 49(5), 724-728.

19. Wierwille, W. W., & Connor, S. A. (1983) Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator. Human Factors, 25(1), 1-16.

20. Hart, S. G., & Hauser, J. R. (1987) Inflight application of three pilot workload measurement techniques. Aviation, Space, and Environmental Medicine, vol. 58, pp. 402-410.

21. Sekiguchi, C., Handa, Y., Gotoh, M., Kurihara, Y., Nagasawa, A., & Kuroda, I. (1978) Evaluation method of mental workload under flight conditions. Aviation, Space, and Environmental Medicine, 49(7), 920-925.

22. Mulder, G., & Mulder-Hajonides van der Meulen, W. R. E. H. (1973) Mental load and the measurement of heart rate variability. Ergonomics, 16(1), 69-83.

23. Ettema, J. H., & Zielhuis, R. L. (1971) Physiological parameters of mental load. Ergonomics, 14(1), 136-144.

24. Opmeer, C. H. J. M. (1973) The information content of successive RR-Interval times in the ECG. Preliminary results using factor analysis and frequency analysis. Ergonomics, 16(1), 105-112.

25. Kalsbeek, J. W. H., & Sykes, R. N. (1967) Objective measurement of mental load. Acta Psychologica, 27, 253-261.

26. Boyce, P. R. (1974) Sinus arrhythmia as a measure of mental load. Ergonomics, 17(2), 177-183.

27. Inomata, O. (1977) An evaluation of heart rate variability in different levels of mental loading. Journal of Human Ergology, 6(2), 208-210.

28. Luczak, H. (1979) Fractioned heart rate variability. Part II : Experiments on superimposition of components of stress. Ergonomics, 22(12), 1315-1323.

29. Coles, M. G. H., Gratton, G., Kramer, A. F., Miller, G. A., and Porges, S. W. (1986) Principles of signal acquisition and analysis. In Psychophysiology : Systems, Processes, and Applications. Vol I : Systems. M. G. H. Coles, E. Donchin, and S. W. Porges (Eds.), New York : Guilford Press.

30. Mulder, G. & Mulder, L. J. M. (1981) Information processing and cardiovascular control. Psychophysiology, 18(4), 392-402.

31. Sayers, B. McA. (1973) Analysis of heart rate variability. Ergonomics, 16(1), 17-32.

32. Sekiguchi, C., Handa, Y., Gotoh, M., Kurihara, Y., Nagasawa, A., & Kuroda, I. (1979) Frequency analysis of heart rate variability during flight conditions. Aviation, Space, and Environmental Medicine, 50(6), 625-634.

33. Egelund, N. (1982) Spectral analysis of heart rate variability as an indicator of driver fatigue. Ergonomics, 25(7), 663-672.

34. Nygaard, B., & Schiotz, I. (1975) Monotoni pa motorveje. Radet for Trafiksikkerhedsforskning, Notat 135, Kobenhaven.

35. Wickens, C. D. (1980) The structure of attentional resources. In R. S. Nickerson (Ed.), Attention and Performance VIII, Hillsdale, N.J.: Erlbaum.

36. Mulder, G. (1979) Sinusarrhythmia and mental work load. In N. Moray (Ed.), Mental Workload: Its Theory and Measurement. New York: Plenum Press.

N88-23386 516-53 140804 5P.

SHORT-TERM MEMORY LOAD AND PRONUNCIATION RATE

Richard Schweickert
and
Cathrin Hayt
Purdue University
West Lafayette, IN 47907

One of the major components determining mental workload is the amount of material that must be maintained in short term memory. Some tasks, such as air traffic control, involve coordination between people, and the main communication is verbal. Critical parts of the communication require memory not only for the gist or meaning of the material, but for verbatum recall (ref. 1). Even tasks which do not involve communication between people often have a verbal component. Communication between humans and computers often requires the human to remember certain information verbatim which has disappeared from the screen (ref 2). Everyone has had the experience of looking up the call number of a book in a library, and rehearsing it while trying to find the shelf.

The capacity of short term memory was described in a classic paper by Miller (ref. 3) as 7 plus or minus 2 chunks, where a chunk is a meaningful unit of material. This gives a good rule of thumb, but it has at least two problems. First, the number seven is an estimate of the memory span, that is, the number of items that can be immediately recalled correctly half the time. But there is nothing special about probability one-half. In most practical situations, we would like to be able to predict probability of correct recall over a range of probabilities, or at least be able to estimate the length of a list that can be recalled with a high probability, say, .99. The second problem is that the probability of correct recall depends on the type of material. The memory span is greater for color names, such as red and orange, than it is for shape names, such as circle and square. Although one can define the capacity of the short term memory to be 7 chunks, this leads to the curious notion that there are more chunks in the name of a shape than in the name of a color.

Another approach is to assume the short term memory is limited in the time for which it can hold items. The support for this has waxed and waned over the years, but the decay hypothesis has enjoyed renewed interest recently. This is because Mackworth, Baddeley, and others have found that the memory span for a type of material can be predicted quite well from the amount of material that can be pronounced in about 1.5 seconds (refs. 4, 5, 6). For example, the memory span for digits is 7.98 and that for four-letter concrete nouns is 5.76 (ref. 7). It turns out that these are the number of digits and nouns, respectively, that a typical subject can pronounce in 1.5 seconds.

This result can be summarized by saying

$$S_i = 1.5 \text{ sec} \times r_i, \qquad (1)$$

where S_i is the memory span for items of type i and r_i is the rate of pronunciation of items of type i, in items/sec.

The explanation is straightforward. Suppose when a subject is presented with material for immediate recall, he forms a verbal trace, and the trace begins to decay. If the subject can emit the items before the trace has deteriorated, recall will be correct, otherwise it will be incorrect. Evidently, on the average, the trace decays after 1.5 seconds, which determines the span.

Equation 1 resolves the second problem, accounting for differences in memory span for different types of material in terms of differences in their pronunciation rates. Schweickert and Boruff (ref. 6) proposed a resolution to the first problem by saying the probability of correct recall is simply the probability that the duration of recall is less than the duration of the verbal memory trace,

$$P = \text{Prob}\,[T_r < T_v], \tag{2}$$

where P is the probability of correct recall, T_r is the time the subject requires to recall the list, and T_v is the duration of the memory trace. In an experiment, subjects were presented with 6 list lengths of 6 types of material. A good account of the data was given by Equation 2. Normal distributions were assumed for T_r and T_v. The mean and variance of the trace duration were estimated to be 1.88 sec and .187 sec^2, respectively.

An equally good, but more easily calculated, estimate of the probability of correct recall was found, based on linear regression,

$$z = -2.02\ T_r + 3.87. \tag{3}$$

Here z is the standard normal deviate of the probability of correct recall of a list, and T_r is the average amount of time required to read the list aloud.

The correlation between the z-score for correct recall and pronunciation time was .977, so 95% of the variance is accounted for by pronunciation time. In contrast, the analogous linear regression equation using the number of items in the list as the predictor yielded a correlation of .849, so only 72% of the variance is accounted for by list length.

It is of interest to note that Equations 2 and 3 underestimated the probability of correct recall for digits, the material subjects had most experience with in daily life, and overestimated the probability of correct recall for nonsense syllables, the material least familiar to the subjects. The subjects in the experiment were not particularly practiced. They came for three one hour sessions, and learned only 60 lists of each material type. The nonsense syllables are hardly chunks, in the usual sense. The following experiment was done to investigate memory in highly practiced subjects.

Method

Subjects. Two subjects completed 4 practice sessions followed by 30 test sessions. They were paid by the hour. Each session lasted about an hour and a half.

Materials. Five types of material were used: consonants, color names, prepositions, shape names, and three letter concrete nouns. To make the

probability of correct guessing low, each set contained 20 items. This precluded the use of digits, a commonly used material in immediate memory studies. Lists of a given material were all presented together in a block. The order of presentation of materials within sessions was governed by six 5 x 5 Latin squares. The lengths of the lists were from 3 to 9 items, inclusive. List lengths were randomized within the blocks.

Procedure. At the beginning of each trial, a list appeared on a TV monitor. In pronunciation trials, subjects read the list aloud with no requirement to remember it. In memory trials, subjects read the list aloud, and then attempted to recall it by speaking aloud. Voice keys indicated the onset and offset of their speaking, and the durations of the utterances were timed with a microcomputer. The pronunciation and recall times are beyond the scope of this paper.

During recall, the experimenter recorded whether the list was correctly recalled or not.

Results

The reading time for a list is the time from when the subject started to read the list until he finished. Reading was followed immediately by recall. Mean reading times and probability of correct recall are given in Tables 1 and 2.

Recall that for the unpracticed subjects in the experiment of Schweickert and Boruff (ref. 6), reading time was a much better predictor of recall than the number of items in the list. Here, the number of items is a better predictor, although only slightly.

For subject 1, the correlation between the z-score for correct recall and the number of items in the list is -.95, so 90% of the variance in recall is accounted for by list length. The correlation between the z-score for correct recall and reading time is -.90, so 80% of the variance in recall is accounted for by reading duration.

For subject 2, the results are similar. The correlation using the number of items in the list was -.95, so 90% of the variance is accounted for by list length. The correlation using reading time is -.92, so 85% of the variance is accounted for by reading time. In each case, list length does slightly better as a predictor than reading time.

The regression equation for predicting the z-score for correct recall is

$$z = b_0 + b_1 n,$$

where n is the number of items. For subject 1, the regression coefficients were $b_0 = 5.50$ and $b_1 = -.83$. For subject 2, they were $b_0 = 5.40$ and $b_1 = -.80$. The coefficients agree remarkably well for the two subjects.

In the calculations, conditions with recall probabilities of 0 or 1 were ignored, since the corresponding z-scores are infinite.

Is there an advantage of practice? One way to evaluate this is to note that the duration of a list recalled half the time was about 2.4 seconds,

compared with 1.8 seconds for the unpracticed subjects in the previous experiment.

Increasing the length of the items leads to two competing tendencies. First, the longer the items, the greater the time required to output the list, so the greater the chances of trace decay before recall is completed. But, second, the longer the items, the more distinctive they tend to be, and hence the greater the chances of guessing an item correctly from a partial trace. Highly practiced subjects are probably better able to reconstruct the partially decayed trace of an item to make a correct guess. The more familiar the items are, the better subjects are able to discriminate the fragments remaining in the traces.

For unpracticed subjects, reading time is a notably better predictor of immediate recall than the number of items in the list. For practiced subjects, the two predictors do about as well, with a slight advantage for the number of items. In either case, about 90% of the variance is accounted for, so for most practical purposes, good estimates of recall probability are available. If the items that must be recalled are likely to be unfamiliar, and likely to remain unfamiliar, then it is advantageous to keep the items short. For example, codes for identifying airplanes or pilots encountered only once in a while should be short to pronounce. On the other hand, if the same items will be encountered over and over again, it is advantageous to concentrate efforts on making them distinctive, even at the cost of adding to the number of syllables.

REFERENCES

1. Lenorovitz, D. R., & Phillips, M. D.: Human factors requirements engineering for air traffic control systems. In G. Salvendy (Ed.), Handbook of human factors. New York: John Wiley, 1986.

2. Card, S. K., Moran, T. P., & Newell, A. The psychology of human-computer interaction. Hillsdale, N. J.: Lawrence Erlbaum, 1983.

3. Miller, G. A. The magical number seven plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63, 81-97, 1956.

4. Mackworth, J. F. The duration of the visual image. Canadian Journal of Psychology, 17, 62-81, 1963.

5. Baddeley, A. D., Thomson, N., & Buchanan, M. Word-length and the structure of short-term memory. Journal of Verbal Learning and Verbal Behavior, 14, 575-589, 1975.

6. Schweickert, R., & Boruff, B. Short-term memory capacity: Magic number or magic spell? Journal of Experimental Psychology: Learning, Memory, and Cognition, 12, 419-425, 1986.

TABLES

Mean Reading Times and Probability of Correct Recall

Table 1: Subject 1

List Length		3	4	5	6	7	8	9
Colors:	Read	.889	1.273	1.656	2.099	2.499	2.945	3.430
	Recall	1.000	1.000	.987	.832	.500	.191	.000
Letters:	Read	.667	.951	1.297	1.659	2.072	2.451	2.896
	Recall	1.000	1.000	.967	.846	.592	.242	.023
Preps:	Read	.867	1.212	1.617	2.018	2.428	2.840	3.275
	Recall	1.000	.993	.940	.805	.415	.113	.020
Shapes:	Read	1.254	1.831	2.399	2.972	3.537	4.037	4.637
	Recall	1.000	.987	.866	.513	.128	.014	.000
Words:	Read	.827	1.195	1.561	1.967	2.380	2.817	3.254
	Recall	1.000	1.000	.931	.685	.281	.055	.000

Table 2: Subject 2

List Length		3	4	5	6	7	8	9
Colors:	Read	.920	1.353	1.764	2.234	2.696	3.151	3.635
	Recall	1.000	1.000	.967	.839	.476	.148	.020
Letters:	Read	.677	1.104	1.468	1.959	2.293	2.743	3.130
	Recall	1.000	.987	.980	.890	.710	.345	.094
Preps:	Read	.883	1.208	1.647	2.067	2.488	2.897	3.287
	Recall	1.000	.993	.967	.879	.537	.208	.053
Shapes:	Read	1.621	2.200	2.790	3.356	3.993	4.574	5.039
	Recall	.993	.953	.800	.547	.157	.013	.000
Words:	Read	.873	1.239	1.664	2.145	2.581	3.010	3.433
	Recall	.993	.993	.927	.627	.366	.088	.007

N88-23387

S17-53
140805
328

ATTENTION, EFFORT, AND FATIGUE: NEUROPSYCHOLOGICAL PERSPECTIVES

Ronald A. Cohen,
Brian F. O'Donnell,

Department of Neurology
University of Massachusetts
Medical Center
Worcester, MA

MK 170541

Attention reflects the deployment of cognitive resources toward internal and external events, and is limited by an individual's capacity for active processing of information. Attentional performance is increased by cognitive effort and diminished by fatigue. Consequently, measures of attention potentially provide excellent indices of ability to resist fatigue and generate effort in the performance of a task. Yet the assessment of attention, effort, and fatigue has been difficult because of problems in operationalizing these phenomena and establishing appropriate methodologies for measurement.

From a neuropsychological perspective, attention is not a single phenomenon, but instead can be dissociated into various components of information processing with biological correlates. Consideration of attention requires a multivariate assessment framework in which variations in stimulus and response characteristics can be simultaneously measured or controlled during serial performance. Attention can then be represented as an index of performance across time. Fatigue may reflect a failure to maintain optimal levels of performance across a number of possible behavioral systems.

Because attention represents multiple processes within the brain and varies over time, the measurement of attention requires similar characteristics. First, attention is a dynamic process necessitating serial assessment in contrast to cross-sectional measurement at a single time point. Second, the assessment should be multivariate, to characterize performance as a function of multiple determinants and outputs. Because attentional processes occur in a biological system, psychophysiological measurement may detect subtle attentional variation based on physiological reactivity.

This paper will review models of attention, effort and fatigue. We will discuss methods for measuring these phenomena from a neuropsychological and psychophysiological perspective. The following methodologies will be included: 1) the autonomic measurement of cognitive effort and quality of encoding, 2) serial assessment approaches to

neuropsychological assessment, and 3) the assessment of subjective reports of fatigue using multidimensional ratings, and their relationship to neurobehavioral measures.

Models of attention

Throughout the history of psychological and cognitive science, attention has been viewed as an important, but difficult to define component of human mental processing. Wilhelm Wundt went to great length in describing the "apperceptive focus" by which an interaction between internal mental events and external reality occurred. (ref. 1)
William James provided the following description:

> " Everyone knows what attention is. It's the taking possession by the mind, in clear vivid form, of one out of what seems several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others."(ref. 2)

Despite early interest in attention, there has been much difficulty in operationalizing or even agreeing to what constitutes the boundaries of meaningful study with respect to attention. In the most narrow sense, attention has been used to describe a nonspecific form of sensory perception. In a broader framework, attention can be viewed as a reflection of the fact that performance is not stable over time. Within this view, attention reflects the temporal variation in performance as a function of both stimulus and response parameters, that are not due to fatigue of peripheral sensory or motor systems.

There are now a number of models that relate attention to both stimulus and response factors. While it is beyond the scope of this paper to review all of the recent models of attention, a review of some of the major positions is in order.

Selective attention

Selective attention refers to the process by which an individual selects stimuli, or components of stimuli from a set of inputs for further cognitive operations. Selection implies a stimulus preference for certain information based on some important feature. Broadbent proposed that selective attention occurs as a function of a limited capacity system in which sensory information is initially processed in parallel until it reaches the level of a filter mechanism. The filter serves to limit the information flow in a serial switching process, so that one input can be processed at a time. This theory predicted that perception was dependent on conscious attention. (ref. 3) Treisman

later demonstrated that certain inputs are potentially perceived even if unattended, though attention was thought to attenuate information and to strengthen the probability of perception.(ref. 4) Both models characterized attention as a sensory mechanism by which filtering of information occurred.

Other investigators offered a much different perspective of attention by suggesting that attention and selection of information occurs at a much later stage of processing. Deutsch and Deutsch proposed that response selection determined selective attention.(ref. 5) The view that attentional selection occurred later in processing was supported by Shiffrin, though the attentional mechanism was proposed to be determined by the rate of transfer and loss of sensory information from short term memory, before loss of information occurred.(ref. 6) The rate of search limited the capacity of attention. An important relationship between memory set size and attentional capacity was noted. Sternberg demonstrated the dynamics of these search processes by exploring whether reaction time in decisions about set inclusion would reflect an exhaustive search of all possible choices, or whether termination of search would occur based on a match with memory.(ref. 7) Initial results seemed to favor an exhaustive search model. However, later investigations suggested that memory set may not define attentional capacity under all conditions.(ref. 8,9) For instance, for well practiced material reaction time for response selection was less dependent on memory set size. These divergent findings led Schneider and Shiffrin to postulate two different attentional processes.(ref. 10)

Automatic versus controlled processing

Within the two process theory attention, a distinction was made between serial "multi-frame" tasks that reflect attentional vigilance, and "single frame" simultaneous display tasks in which accuracy of perception is high, but reaction time is affected by search requirements. The two processes were suggested to differ on the dimension of controlled search versus automatic detection.

The paradigm developed for study of the two process model of attention (ref. 10) manipulated a number of important variables that affect attention. The independent variables included: frame time, memory set, frame size, and type of spatial mapping. Frame time reflected the speed of stimulus presentation. Increased speed obviously increased attentional demand. Memory set referred to material presented to the subject prior to the search task, which determines the familiarity with the stimuli during attention. Frame size determines the number of stimuli that have to be searched before a decision can be made about the

presence of a target stimulus from the memory set. The type of spatial mapping was either consistent or variable. In the consistent mapping task, subjects always searched for a given stimulus type among a set of consistent choices. In the variable mapping condition, a random subset of target stimuli were presented on each trial, so that subjects could not anticipate the upcoming stimulus. Within the multiframe search paradigm, a number of trials were presented with the task of accurate detection of target stimuli. Accuracy was strongly related to memory set and frame size in the variable mapping condition, but not in the consistent mapping condition. Under consistent mapping condition, perceptual factors played a larger role. Also, mean reaction time increased in a linear fashion as a function of memory load and frame size.

Interestingly, mean reaction time was not the only factor determining the role of memory and frame size, as variablity of reaction time was greatly affected by load when a consistent memory set was not present. Attention was shown to be dependent on the extent of familiarity or previous practice with the information to be processed on consistent mapping tasks, but not with variable mapping, suggesting that with variable mapping, familiarity is much more difficult to obtain and automaticity is not possible. When consistent mapping was used, even complex information could be processed "automatically" with little attentional capacity allocated to them under conditions of high familiarity. However, under novel conditions greater task demand exists and a more effortful, sequential type of processing is required, making automaticity impossible. There is now indication that this type of processing loads more on the response system.(ref. 11,12)

The paradigms suggested in the two process model are important since they establish critical parameters for consideration of attention. To summarize, these parameters include variables such as frame size and complexity, demands on memory set, the consistency of stimuli to be attended, as well as the nature of stimulus presentation(e.g., multi-frame or single-frame). Control of these variables is necessary for the adequate study and measurement of attentional variation. Furthermore, an implicit component of this paradigm is that perceptual requirements are within certain boundaries. When the perceptual task is more complex, there may be a greater effect on attention. Table I defines some of the relevant variables to be manipulated in the measurement of attention.

The spatial characteristics of directed visual attention have been studied using paradigms involving automatic search strategies. For instance, Hughes and Zimba used spatial precuing techniques to prompt attention on a signal detection task (ref. 13). Using reaction time

measurement, spatial maps were created based on expectancy effects. Attentional expectancy seemed to be greatest along the major vertical and horizontal meridians suggesting a spatial gradient that may exist along intrinsic cartesian coordinates. Other recent studies have shown similar topographic maps of three dimensional space that relate attentional expectancy with spatial location.(ref. 14) Correct directing of attention to spatial location can have small, but significant benefits, while incorrect orientation can have significant costs. The cost of inattention increased as a function of spatial distance between cued location and actual target position. The reorientation of attention across the horizontal and vertical meridians of space was related more to ocular-motor requirements for search rather than pure sensory mechanisms. The implications of these studies, as well as recent work in other primates(ref. 15), is for an increased role of motor and pre-motor influences in directed attention.

Attentional capacity

Kahneman suggested that there is only a limited degree of energy available for performing mental operations which limits the capacity of all stages of information processing.(ref. 16) Since the earlier stages of sensory processing occur more "automatically", attentional demand is not as great as when response production is required. An important aspect of this model is the emphasis placed on the energetics of the individual's nervous system in defining capacity.

Shiffrin and several of his colleagues have investigated the parameters that determine capacity and have demonstrated equal successive and sequential performance under certain conditions of consistent mapping.(ref. 17. 18, 19). This finding was initially viewed as a reflection of parallel processing and indication that capacity was not important. However, subsequent interpretations have suggested the conditions of consistent mapping and fixed stimuli across trials may lead to conditions of automaticity. With automaticity, parallel processing is possible and capacity is less relevant. However, under conditions of variable mapping, when such processing is not possible, sequential processing is required and is determined by rate limitations of the individual's capacity.

While defining what constitutes "available energy" is difficult, the capacity model creates an important bridge to the biological characteristics of the individual. This biological constraint consists of a natural variation over time as the state of the organism changes. Furthermore, this approach predicts a relationship between attention and both central and peripheral nervous system arousal and activation. Attention cannot be viewed solely as an

information processing system, since important neurobehavioral constraints exist which mediate attention. Neurobehavioral factors may ultimately define the salience of stimuli which serve to drive attentional states.

Consideration of most current attentional models leads to a reciprocal linkage between effort and task characteristics. Tasks with high demand characteristics require greater attentional effort for successful performance. Salient stimuli or tasks with less inherent demand may elicit greater attentional effort as a result of their meaningfulness or contextual relevance. An example of the effects of salience on attention and effort may be seen in other paradigms ranging from early works on classical conditioning and the orienting response(ref. 20) to more recent cognitive studies of memory encoding. In the case of the orienting response, the salience of a stimulus in the environment ultimately determines the intensity of attentional response that occurs. This relationship will be discussed in greater detail latter in this review.

Memory and attention

Recent studies of memory encoding have also suggested a relationship between task salience, the amount of attentional effort directed on the task, and incidental memory performance. Jennings and his colleagues demonstrated a relationship between variations in alertness and attention during encoding of semantic information, and subsequent recall of the information.(ref. 21) Cohen and Waters demonstrated a similar effect using a levels of processing memory paradigm.(ref. 22) When subjects performed more salient semantic tasks, as compared to phonemic and less salient semantic tasks, greater attentional activation was noted. This finding seemed to suggest an attentional explanation for the levels of processing effect, rather than the usual interpretation of a spread of activation or associative elaboration process for semantic information. While the levels of processing effect may have occurred because of greater associative elaboration on semantic tasks, it was clear that the semantic task also elicited a greater attentional response on trials that were subsequently recalled. These studies suggest that in information processing tasks requiring complex mental operations, there is substantial attentional variation that is intimately linked to the salience of the information at each moment of presentation. Furthermore, physiological activation was highly correlated with this attentional effect.

The linkage of memory encoding effects with selective attention is extremely important at a conceptual level. Typically, attention is studied as an isolated function

within paradigms that are designed to demonstrate relevant components. However, studies of other cognitive functions such as memory encoding often demonstrate strong attentional bases for memory effects. While there is a danger in making attention over inclusive, it is clear that there are attentional components in literally all cognitive tasks. Therefore, the methodological chore of researchers in this area is to establish techniques for extracting the attentional component from a range of cognitive tasks or formal neuropsychological measures. By doing so it may be possible to determine the consistent capacity that an individual possesses for a given cognitive function(e.g., word generation) from the alterations in performance that are associated with variations in attention. Based on the large body of literature previously mentioned, this variation in attention will be dependent on numerous factors including: spatial topography of the information, memory load, rate and redundancy of presentation (ie., temporal characteristics), salience of information, the modality of the task (e.g., language vs. visual analysis), as well as the available capacity or "energy" for the task.

Influence of fatigue

The term fatigue has been even more problematic for investigators than that of attention and effort. Part of the difficulty in studying fatigue has arisen from fundamental disagreements on relevant systems of study, as well as the proper level of analysis. Physiologists have long referred to fatigue within the context of neuromuscular changes occurring at a peripheral or even cellular level. On the other hand, behavioral investigators use the term fatigue to refer to subjective experiences of difficulty or inability to persist on tasks. Fatigue may also refer to actual performance decrements over time that are related to central processing depletion. In this context fatigue may have some relationship to the process of habituation, though it is typically assumed to involve a failure of action, rather than a passive extinction.

Because of the many ways in which fatigue can be conceptualized and defined, some theoreticians have questioned the usefulness of the term.(ref. 23) Broadbent discussed the difficulties in developing a test of fatigue and also raised questions about the utility of the construct.(ref. 24) Nevertheless, fatigue is a reported experience of individuals under conditions of prolonged effort, task demands and vigilance. To some extent the experience of fatigue bears on questions related to processing capacity. Clearly, changes in central nervous system arousal may elicit this experience even in the absence of a task. This is commonly noted in individuals with affective disorders,(ref. 25) as well as certain

neurological disorders affecting subcortical systems. (refs. 26, 27)

An important differentiation must be made in the study of fatigue between the subjective reports regarding an individual's experience of fatigue and actual performance decrements. These two factors may or may not be linked. A description of some recent developments in the study and measurement of fatigue will be discussed later in this paper.

Physiological correlates of attention and effort

Early studies of autonomic psychophysiology related the orienting response to attentional registration of novel stimuli. Sokolov placed much emphasis on defining the orienting response relative to perceptual matching of incoming stimuli with existing "neuronal models".(ref. 28) The diminishing of the orienting response over repetitive nonreinforced trials was defined as habituation. Much research has focused on determining whether the process of habituation is a passive extinction of response, or whether it requires an active neuronal mechanism that overrides the attentional response of orientation and causes inhibition.(ref. 29) While various neurophysiological mechanisms have been identified that may underlie orientation and habituation(ref. 30), the integration of these fundamental psychophysiological processes into more complex cognitive paradigms has been more difficult.

Recent investigations have revealed that autonomic reactivity is differentially associated with a variety of factors ranging from sensory factors to "cognitive load" and the amount of memory involvement. Lacey and Lacey provided one of the first formulations of a differentiation of heart rate response related to different task demands.(ref. 31) Heart rate deceleration was thought to relate to environmental intake, while acceleration was associated with cognitive elaboration or the rejection of information. A number of subsequent investigations examined these relationships, though more emphasis was initially placed on defining the role of cardiac deceleration in passive attention. A number of characteristics that influence this response have been investigated and for the most part they reflect the constraints of the orienting response as described by Lynn and others. Important characteristics include: 1.Stimulus significance , 2.Expectancy, 3.Stimulus intensity and rate of onset, 4.Estimation of stimulus contiguity, 5.Termination of anticipation, 6.Perceptual factors, and 7.Stimulus detection difficulties(i.e., noise or interference).(ref. 20, 32, 33) Much debate has centered on the specific role of cardiac deceleration, with explanations ranging from "the holding of available

processing capacity"(ref 34) to an enhancement mechanism for perceptual processing.(ref. 35) Obrist and his colleagues have argued that deceleration occurs due to "motor quieting" in readiness for response.(ref. 36) Without attempting to resolve this debate at the moment, it is evident that most explanations propose an adaptive basis for this response. Cardiac deceleration seems to be related to the readiness for perceptual intake.

A few studies have investigated the role of cardiac acceleration in information processing. Kahneman et al. showed a direct relationship between acceleration and transformation difficulty on a paced serial addition task.(ref. 37) Jennings later showed a dissociation between attentive listening which has been shown to cause deceleration and cardiac acceleration associated with cognitive transformations when input and output requirements were controlled.(ref. 38) This relationship has been expanded in several studies in which physiological reactivity was shown to vary in subtle ways based on conditions of memory encoding. Jennings and Hall varied memory load on task in which subjects were to process 5 and 10 item word sets.(ref. 21) Cardiac acceleration was related to the encoding phase of the task, in so much as words later recalled had greater acceleration associated with them. However, acceleration was not related to cognitive load (ie., number of items in set) directly. Instead increased acceleration occurred during encoding and seemed to relate to the degree of directed attention or effort. This interpretation was also offered in other studies. (ref. 39, 40)

Attention has been studied in other physiological systems. Kahneman and Beatty demonstrated a direct relationship between pupil dilation and the amount of information being processed in short term memory.(ref. 41) Siddle and his colleagues found that skin conductance response occurred proportional to the degree of shift in semantic category.(ref. 42) Yuille and Hare extended these findings to include a variety of other autonomic measures and also showed a direct relationship between short term memory and physiological reactivity.(ref. 43)

The results of most previous studies of information processing components and physiological reactivity suggested a relationship between task demands and autonomic activation. Except under certain cases of perceptual intake, cardiac acceleration occurred as tasks required increasing cognitive manipulations. Furthermore, a relationship between physiological activation, task demands and ultimately short term memory characteristics emerged.

Analysis of the motor systems has revealed interesting relationships between motor activation and

cognitive processes. The history of theories of covert motor involvement in thinking has origins in the work of Watson on subvocalization. Recent studies by Cacioppo and Petty have suggested that cognitive tasks such as classification based on attributes may result in different patterns of muscle response (EMG) depending on the required "level of processing".(ref. 44) Other investigators have related the memorization of words to EMG changes based on the phonetic features, thus providing evidence for a subvocal motor mechanism.(ref. 45) However, studies investigating the effect of suppression of subvocalization on short term memory performance have provided mixed results.(ref. 46)

Cohen and Waters (ref. 22) provided a methodology for dissociating some of the effects of motor system activation from other autonomic correlates of memory performance. This study provides a good example of how physiological measures can help to identify important factors that are not apparent from behavioral or cognitive measures alone. For instance, the levels of processing effect was initially viewed as a function of the "depth" or elaboration of associative processes. Cohen and Waters provided evidence for the role of attentional and psychophysiological activation in mediating the levels effect. This activation would not have been apparent without the use of physiological measurement. Since this study is an illustration of the merging of psychophysiological methods with paradigms derived from cognitive psychology, further discussion of the specific methodology and results will be provided.

Levels and stages of processing.

Cohen and Waters (ref. 22) demonstrated a levels of processing effect(ref. 47) in which words processed using more complex semantic operations resulted in greater incidental memory than words processed with less complex semantic operations. Both semantic tasks produced better recall than a phonemic task. Within this memory framework a paradigm was created that allowed for measurement of several physiological systems during different stages of the task. Figure 1 contains a schematic diagram of the stages of processing. Subjects first were presented with a cue stimulus that identified the required level of processing for the upcoming word. Seven and a half seconds later a word appeared, and subjects were required to covertly think of responses for the word (7.5 sec). In the third stage, subjects were asked to vocalize their response during a twenty second interval. Thus, analysis of heart rate, skin conductance, skin temperature, and two sites of EMG was conducted across the three stages of processing. Dependent measures consisted of the phasic change for each physiological measure relative to baseline, determined by

subtracting the average activity during a rest interval prior to each trial from the activity during each stage of the processing for that trial.

A number of interesting results emerged from the study. First, across all three levels of processing there were significant increases in physiological activation for each word item for the later stages of the task(ie., vocalization produced more activation, as compared to covert processing, and covert processing produced more activation than the cue/anticipation stage). This activation was significant across a number of systems including heart rate, skin conductance and EMG. The magnitude of response change was greatest for skin conductance and heart rate. The heart rate response always reflected an acceleration, even in the case of the cue/anticipation stage, suggesting that even though this stage involved readiness for a stimulus, the anticipatory effect was reflected in the cardiovascular system. This finding suggested that under conditions of increased task demand the deceleration response may be overridden by competing factors of arousal associated with expectancy. This anticipatory response habituated over the course of the 39 trials, while response to covert processing and verbalization failed to habituate.

With respect to the levels of processing effect, a different pattern emerged. Significant effects as a function of level of processing were noted for heart rate and skin conductance, but not for EMG. Furthermore, there was a significant interaction between level and stage of processing, such that the verbalization of a response tended to be the point in which the greatest differential activation across levels occurred. Yet, this activation was not related to the overt motor demands of verbalization, since a levels effect was not noted in the EMG system. Table 2 shows the relationship between levels and stage of processing across the different physiological systems. These findings indicate an important relationship between task characteristics during encoding, the production of actual responses, and physiological activation. Figures 2 and 3 illustrate the main effects of this study.

As Jennings and Hall had noted earlier, retrospective comparison of items that were later recalled with those that were not, suggested that greater heart rate and skin conductance responses occurred during the encoding stages of those that were later recalled. In addition, the degree of activation noted on recalled trials was unrelated to the levels of processing effect. Therefore, regardless of the task type(i.e., semantic or phonemic), if physiological reactivity was greater on a particular trial, the information was more likely to be recalled later. This finding illustrates the close link between physiological activation and the encoding process and goes

further to point to the role of attentional direction and effort in the production of successful processing of information. The findings also indicate that greater the response production requirements during later processing stages results in greater physiological activation, illustrating the importance of response factors in this type of attentional task.

The significance of the recent studies of physiological correlates of attention are twofold. First, these studies point to relationships between components of attention and physiological activation. Clearly, as task demand increases there is an increased effect on later response components of attention. Secondly, these studies illustrate attentional influences in a variety of tasks that may normally be interpreted as being outside the realm of domain of attentional consideration. Therefore, the study of physiological correlates of cognitive performance provides an important methodology for obtaining indices of attentional variation during performance. These studies emphasize the importance of viewing attention and effort from a biobehavioral standpoint with possible implications for the adaptive functioning of the individual.

Neuropsychological Measurement and Attention

The study of brain-behavior relationships has made significant progress, in part because of refinements in assessment methodologies for accurately measuring and quantifying cognitive performance. One of the foundations of neuropsychological assessment is the use of multivariate approaches that allow for a broad cross section of many different cognitive functions. By comparing an individuals performance across these functions to established normative data, it is often possible to provide detailed information about deficit patterns. Pattern analysis of cognitive deficits can give evidence of localized brain dysfunction. The multivariate approach is critical to neuropsychological assessment because it allows for analysis of common variance across many different measures, as well as the unique variance associated with a particular measure or behavioral deficit.

Neuropsychology has been very successful in identifying performance deficits that may correlate with structural brain dysfunction. An individual's performance can be mapped across areas encompassing language, visual perception and integration, memory, motor dexterity and executive response capability. A mosaic of data emerges from the assessment that provides a cross-sectional picture of the patient's abilities. Neuropsychology has been particularly effective at measuring and providing anatomic mapping of the more static functions such as language and visual perceptual

performance. The dynamic functions of memory and executive control have also been addressed within neuropsychological methodologies. In the case of memory assessment there has traditionally been an emphasis on intentional learning paradigms, though there has been an increasing inclusion of paradigms that assess other memory modalities(e.g., episodic memory). However, the dynamic models that integrate and assess attentional variation have been under utilized in neuropsychology. A reason for this lack of inclusion of attentional methodologies may stem from inherent difficulties in modifying existing tasks to account for attentional effects. Also, the problems associated with the operationalizing of attention (as mentioned earlier) may also account for this neglect.

Within current neuropsychological methodologies, attention is either addressed through interpretation of behavioral tendencies cutting across all tasks of the assessment process, or through certain tests that are thought to load primarily on attentional factors.(For a general review of approaches to neuropsychology and assessment see ref. 48, 49, 50) In the case of the first approach, the clinician usually makes a judgment about the presence of attention deficits based on behavioral observation of the patient's response tendencies. For instance, a patient who clearly shows the capacity to perform certain types of tasks, but who doesn't perform in a consistent manner is often described as showing attentional variation. Many of these behavioral observation approaches have been formalized in the assessment of children with attention deficit syndromes. The work on attention deficit disorders of children has yielded some of the strongest methods for assessing attentional variation, which has led to success in quantifying the degree of deficit in attention, as well as subtypes of attention deficit disorders in children.(51, 52) The methodologies used in this area also have applications for assessment of adult variation in performance.

Several standardized tests are commonly used to assess attentional performance. Since the list of these measures is rather short, a brief description of the most generally used of these tests will be given.(see ref. 48) These include the digit span test, the trail making test, the cancellation tasks, the paced auditory serial addition test(PASAT), the symbol digit modality test, the Stroop test, the continuous performance test, and the span of apprehension tasks.

Within the context of intellectual assessment, the use of selective subtest patterns to define potential deficits has been well established. Specific deficits on the digit span, arithmetic and digit symbol subtests have been associated with certain attention deficit disorders. The

digit span backwards task seems to require considerable mental control and effort. As a result it is very sensitive to deficits affecting memory and response control, though certainly a variety of factors may result in poor performance on this task. The strength of this task lies in the fact that it contains a gradient of difficulty that increases effortful demands as digit length increases. However, on a given trial the time required for sustained vigilance is short, so that assessment of serial attentional variation is difficult. Thus, this task seems to measure short term memory and the capacity for effortful manipulations over a few seconds.

Both the digit symbol and symbol digit modality tasks are among the most sensitive neuropsychological measures for detecting brain impairment since a host of factors can result in impaired performance. Disorders affecting arousal, as well as motor speed will clearly cause problems. Also, task performance will be negatively affected by memory limitations, encoding difficulties or even visual perceptual deficits. However, in the absence of deficits in these other functional areas, it may be safe to assume that performance on this task reflects the ability to maintain consistent and rapid performance for longer intervals under conditions of high cognitive load for new information.

The trail making test may be more analogous to Shiffrin and Schneider's "single frame" method since visual search is required on a fixed map containing twenty five points. However, unlike that paradigm a physical response of tracking between points is required. Since the sequence of points to be tracked on Trail A contains a continuous series of numbers, the task should be rather automatic with respect to memory and cognitive demand, and most of the task's effort is associated with visual search. On Trail B, a more complex task is given in that subjects must alternate between a number and letter sequence of twenty five items. The demand of this task is greater since a switching operation is required in conjunction with visual search. This added demand typically results in a slowed response time, even in normal individuals. This task probably is affected by attentional variation, though in its normal administration, the only dependent measure is total time for completion of the sequence. Therefore, attentional variation during the course of the task is not to be measured.

Vigilance within a neuropsychological framework usually refers to the ability to sustain attention for prolonged time periods. Tasks which assess vigilance are typically simple to perform on a single trial in that detection of a stimulus (e.g., number or letter) is required. Difficulty arises because of the large quantity of trials that are administered, and the fact that it is often taxing to persist over time. The strength of this

type of task is that it fits the multiframe methodology with low memory load and relative ease of detection for any given trial. However, because these tasks are inherently simple in stimulus and response characteristics, vigilance tasks lack salience or contextual relevance. While they do provide a relatively clean measure of sustained attention, they may reflect more on capacity to resist habituation or "boredom" rather than to actively process the environment. The most commonly used measures of vigilance include: the letter, digit and symbol cancellation tasks, the perceptual speed task, and the continuous performance task. The cancellation tasks are among the most easily administered vigilance tasks, and are easier than the perceptual speed and continuous performance tasks, in that stimuli and response times are not varied. On the perceptual speed task, the target stimulus shifts between lines, thereby requiring a shift in response set. The difficulty of the continuous performance task arises from the speed of stimulus presentation and the requirement that the subject keep up response speed as well as accuracy.

The Paced Auditory Serial Addition Test is a good multi-frame task of attention. On this test a continuous performance format is used in which the rate of stimulus presentation is controlled. Instead of a simple detection task, this test requires subjects to add a number that is being presented to a number which has been previously presented. The strength of the test stems from the use of a more complex cognitive operation in conjunction with a methodology requiring sustained performance. Since the cognitive operation (addition) is relatively easy, the effect of task complexity is mainly to increase the required effort and attentional demand. However, this demand is controlled for by the rate of stimulus presentation. Faster presentation results in an increased difficulty level. Therefore, this test controls for many of the requirements for multi-frame tasks as described by Schneider and Shiffrin, though the mental operation is more complex.

The Stroop Test measures a somewhat different attentional component, associated with focused attention and freedom from distractibility. On the interference trial, subjects must block the effects of non-relevant information. The task requires an override of automatic processes of word reading for successful completion and therefore has interesting theoretical implications. The Stroop test shows that automatic processes can be countered, but that this requires much effort and that capacity is affected. Many norms are available on this test for a variety of different clinical populations.

There are several other measures included under the category of measures of executive control that reflect on

motor, pre-motor and response control and planning capabilities. These include the motor impersistance tasks, the grooved pegboard task, the word generation task, and the Porteus Maze Test. All of these measures have strong attentional requirements. For instance, the Stroop Test directly measures freedom from distractability or interference. It is beyond the scope of this paper to review all of these measures; however, it should be noted that measures of executive control emphasize the response system's capabilities.

Review of the set of "attentional" measures used in neuropsychological assessment leads to a mixed appraisal. On one hand, a number of strong measures exist that require simple cognitive operations, and therefore allow for an index of the ability to persist on multiframe serial tasks, and on single frame search tasks (e.g., Trail Making). Analysis of performance on these measures may lead to isolation of different types of attentional difficulties, especially in the absence of deficits on other more "static" measures of cognitive function.

The major limitation of these tests results from the very nature of these attentional tasks. Many of the tasks are designed to make minimal demands on cognitive operations. Therefore, these tests generally fail to measure attentional variation associated with the complex types of information processing required in many situations. Individuals may be capable of persisting on simple vigilance tasks, but may show considerable attentional variation on more complex tasks, or in certain modalities of function (e.g., the processing of language input). Therefore, there is a need for the development or modification of cognitive tasks that will allow for assessment of serial variation in performance across a variety of functional areas.

Methodological adaptations and applications

Most approaches to neuropsychological assessment are multivariate and therefore meet an important requirement for enhancing current methodologies for measuring attention. As mentioned previously, a fundamental problem with the current attentional measurement systems results from a tendency to measure attention in the context of a special attentional test. By design, most attentional tests control for task demands, stimulus characteristics and other variables so as to directly measure vigilance, visual search or some other attentional component. However, these tasks often lack contextual relevance and do not allow for an understanding of attentional variation in the course of performing naturalistic cognitive operations. While performance of addition during paced serial presentation

does reflect on the effects of increased cognitive load over time (ie., persistence), this type of task fails to account for how attention varies under more contextual demands or when other specific cognitive functions are required.

To address this problem, Cohen, O'Donnell and several colleagues are developing methods for measuring the variation in test performance accounted for by attentional fluctuation. The general approach to this work has been the modification of existing neuropsychological measures so as to allow for assessment of serial variability. In addition, several existing measures are analyzed with respect to the degree of consistency in performance. Examples of these modifications are reflected in the analysis of performance on measures such as Digit Span or the Peterson Distractor Task. In the case of Digit Span, analysis of the number of trials in which one of the two digit sequences is missed is conducted. One would expect that if the errors were due primarily to the length of the digit sequence, error should occur only on the last one or two sequence lengths. If attentional variation related to some other influence is playing a role, errors may be expected on earlier trials. In a sense, this method formalizes a method of interpretation that is often anecdotally described and used by clinicians. In the case of the Peterson Distractor Test, a modification was made so that a repeated measures analysis could be performed, thus indicating whether there was significant variability across trials of the test, independent of the length of the distractor period for a given trial. While analysis of recall as a function of time of distraction may be more important from the perspective of memory assessment, a repeated measures analysis may reveal more information about attentional variation.

Similar modifications were made in a variety of other neuropsychological measures to allow for either repeated measure analysis of performance across trials or an analysis of performance across blocks of time when a particular test is not structured with trials. The other tests in which these modifications were made include: Word Generation, Grooved Pegboard, Symbol Digit Modality Test, Trail Making Tasks, Stroop Interference Test, and the Continuous Performance Test. These tests were analyzed in manner similar to that described for the Peterson Distractor Task.

Several standard measures of cognitive functioning were modified in a way more analogous to that described for the Digit Span Test, in that analyses were conducted to determine the degree of variability in performance. For instance, the Paired Associate Learning Test was analyzed to give the number of instances when a drop in performance was

noted over successive trials. Since a normal learning curve would be expected, such a drop would suggest attentional variation. Similar strategies for analysis were used of the Block Design subtest and other verbal subtests of the WAIS-R.

These methods were later applied to investigations of patients with affective disorders, multiple sclerosis and other neuropsychological damage. The conclusion of this paper will propose applications of our neuropsychological and psychophysiological methodologies to the assessment and prediction of attentional variation and performance. Examples from recent clinical investigations will be discussed to demonstrate the sensitivity of these techniques.

Fatigue associated with multiple sclerosis.

While fatigue is not the most debilitating problem associated with multiple sclerosis (MS) it has been shown to be the most frequently reported symptom of MS patients and may be the most frustrating for some individuals.(ref. 27) However, there has been disagreement on whether symptoms of fatigue in MS are associated with actual neuromuscular deficits, or with central nervous system effects associated with motivational or cognitive changes. Furthermore, the relationship between subjective complaints of fatigue and behavioral decrements was not clear.

Cohen and Fisher recently studied 29 patients with MS whose illness was of moderate severity.(ref. 53) Patients were assessed using many of the measures mentioned in the preceding section. Evaluations were repeated, so that each patient had three evaluations in conjunction with a cross over design drug study (Amantadine). While almost all patients showed substantial motor slowness on all measures, this was not the deficit most often associated with the symptom of fatigue as reported by the patients. Instead, it was noted that patients showing greater within test variability were more apt to report fatigue. Fatigue seemed to be associated with capacity to sustain consistent performance. Patients with greater variability in general performance also tended to have more difficulty on memory tasks, particularly when distraction was involved. Fatigue consisted of an increased variability in performance, rather than a linear decrement over time.

Patients maintained a fatigue diary that was scored using a multi-dimensional rating system. Based on subjective reports, fatigue was usually felt to involve motivational and general changes in "energy" rather than muscular weakness or tiredness.

The study of fatigue in MS may be of broad interest

because of the role of subcortical influences in mediating motivational and energy states. MS affects the white matter of the lower brain systems and presumably disrupts the ease of signal transmission in the brain. Therefore, the present findings illustrate the relationship between neurological systems involved in arousal and nerve signal transmission, and associated cognitive, as well as affective changes. From a methodological standpoint, this study demonstrates the importance of serial assessment approaches, since variability across trials proved to be the most important correlate of the attentional difficulties and fatigue noted by patients.

Attentional bases of affective disorders

In another investigation, Cohen, Fennell and Bauer investigated two groups of patients with major affective disorders.(ref. 54) One group of 19 patients were diagnosed as manic, while a second group of 24 patients were experiencing major depression with symptoms of psychomotor retardation. Comprehensive neuropsychological evaluations were conducted on all patients. In addition, several of the manic patients were studied longitudinally to provide indices of any fluctuation in performance as a function of changes in their bipolar state. Previous neuropsychological investigations had suggested that non-dominant hemisphere dysfunction was a correlate of affective disorders. This interpretation was based on the finding of greater impairment on non-verbal visual motor tasks, as well as other tasks thought to be associated with the right hemisphere.(refs. 55, 56) However, analysis of performance on many of the "non-dominant hemisphere" tasks revealed that the difficulties of the affective patients was often due to a failure to generate sufficient effort and to maintain consistent attention. Tasks such as the Block Design subtest require much greater effort for successful completion than many verbal measures such as the Vocabulary subtest. On many of the verbal tests, performance is determined by either the presence or absence of a certain competency level. On many non-verbal tasks, competency may present and yet scores will be low if an individual does not have the capacity to persist. It is not surprising then that patients with disorders affecting motivation, energy level and drive would have greater difficulty on tasks with greater demands. Analysis of bipolar patients during different stages of affective state reveals variation in error types depending if they are manic or depressed.

The study of attentional variation in affective disorders illustrates the importance of extending performance measures to give indications of temporal variability in performance. The importance of determining the task demands and their relationship to available capacity is also evident.

Circadian and Drive State Influences

The performance of individuals under altered states of arousal and affect was noted in the affective disorders, and is in even more apparent in patients with damage to subcortical systems that control arousal and drive state. The importance of maintaining an optimal state of arousal during information processing was alluded to before in discussion of psychophsyiological mediation of attention. Kleinsmith and Kaplan demonstrated an inverted U shaped function with optimal memory performance during periods of moderate arousal.(ref. 57) More recently, studies by Folkhard and Monk have demonstrated that this effect operates under circadian influences, and that optimal memory performance tends to occur at certain times of the day.(ref. 58, 59)

Cohen and Albers recently studied a woman (AH) with a history of craniopharyngioma that had adhered to the inferior hypothalamus.(ref. 60) AH presented with a strikingly impaired sleep-wake pattern, as well as other disturbances of arousal and behavior. The basis of her behavioral disregulation seemed largely related to destruction of the suprachiasmic areas of the hypothalamus, which in other animals has been shown to serve as a circadian pacemaker. The unusual aspect of AH's presentation was the extreme irregularity and variation in behavior from moment to moment. Figure 4 is a graph of her sleep-wake cycle over a week's time. Neuropsychological studies conducted over the course of three separate sessions revealed similar inconsistency in cognitive performance (see Table 3). Even within the course of an hour during the evaluation the patient exhibited moments of excellent performance (e.g., successful completion of the most difficult items of the Similarities subtest) that would be followed by periods of complete logical failure. In this case, scores on a particular test do not provide an index of the type of dysfunction that had occurred. Only through longitudinal analysis of performance over the course of the session and across multiple sessions, could the fluctuation in arousal and executive control be determined. This case study again demonstrates the delicate balance between internal systems regulating drive states, arousal and mental processes, and the expression of behavioral disturbances of attention and arousal.

Work site applications

While neuropsychology usually emphasizes the study of cognitive functions in brain injured individuals, this methodology may be easily adapted for assessing and predicting normal human performance. The neuropsychological evaluation we propose utilizes three

types of measures to concurrently assess attentional variation: 1.Performance, 2.Physiological response, and 3.Subjective report. The adaptation of the methodologies discussed earlier in this paper primarily requires a situation that allows for concurrent sampling of physiological and performance variables over the course of normal work activities. Table 3 lists some of the variables that need to be considered.

The aerospace setting should be ideal for this, since physiological measurement may already be part of the operating procedure. After sampling relevant variables, correlation of physiological measures with behavioral and cognitive responses would allow for a determination of indices that reflect attentional variation. The determination of specific markers in behavioral and physiological response that predict impending attentional failure, would be the ultimate goal this strategy. Also, the use of subjective measures of self-report of fatigue and mood may provide important information, though there still needs to be much work in determining the relationship to actual performance decrements.

The use of neuropsychological methods provides a useful approach to the assessment issues in the work site. These tests have been shown to be very sensitive to changes in different cognitive functions, and they reflect fluctuations in brain state. There is also a large body of normative data for most common neuropsychological tests. As mentioned previously, adaptations can be made to make various tests more sensitive to attentional demands, and to enable an extraction of variance associated with attentional fluctuation.

REFERENCES

1. Pillsbury, W.B. Attention. Chapter 1. Macmillian and Co., NY, 1908.

2. James, W. The Principles of Psychology. Chapter 11. Attention. Henry Holt and Co., NY, 1890.

3. Broadbent. D.E. Perception and Communication. London: Pergamon, 1958.

4. Treisman, A.M. Strategies and Models of Selective Attention. Psychological Review, 1969, Vol. 76, 282-299.

5. Deutsch, J.A. & Deutsch, D. Attention: Some Theoretical Considerations. Psychological Review, 1963, Vol. 70, 80-90.

6. Shiffrin, R.M. The Locus and Role Of Attention in Memory Systems. In P.M.A. Rabbitt & S. Dornic (Eds), Attention and Performance V. New York: Academic Press, 1975.

7. Sternberg, S. High Speed Scanning in Human Memory. Science, 1966, Vol. 153, 652-654.

8. Simpson, P.J. High Speed Scanning: Stability and Generality. Journal of Experimental Psychology, 1972, Vol. 96, 239-246.

9. Briggs, G.E. & Johnsen, A.M. On the Nature of Central Processes in Choice Reactions. Memory & Cognition, 1973, Vol. 1, 91-100.

10. Schneider, W. & Shiffrin, R.M. Psychological Review, Vol. 84, 1977, 1-66.

11. Fischer, B. & Breitmeyer, B. Mechanisms of Visual Attention Revealed by Saccadic Eye Movements. Neuropsychologia, 1987, Vol. 25, 73-83.

12. Posner, M.I., Snyder, C.R.R. & Davidson, B.J. Attention and the Detection of Signals. J. Exp. Psychol., 1980, Gen. 109, 160-174.

13. Hughes, H.C. & Zimba, L.D. Natural Boundaries for the Spatial Spread of Visual Attention. Neuropsychologia, 1987, Vol. 25, 5-18.

14. Gawryszewski, L.R., Rizzolatti, G. & Umilta, G. Movement of Attention in the Three Spatial Dimensions and the Meaning of Neutral Cues. Neuropsychologia, 1987, Vol. 25, 19-29.

15. Goldberg, M.E. & Seagraves, M.A. Visuospatial and Motor Attention in the Monkey. Neuropsychologia, 1987, Vol. 25, 107-118.

16. Kahneman, D. Attention and Effort. Englewood Cliffs, N.J., Prentice-Hall, 1973.

17. Shiffrin, R.M., Gardner, G.T. & Allmeyer, D.H. On the Degree of Attention and Capacity Limitations in Visual Processing. Perception & Psychophysics, 1973, Vol. 14, 231-236.

18. Shiffrin, R.M., McKay, D.P. & Shaffer, W.O. Attending to Forty-nine spatial positions at Once. Journal of Experimental Psychology: Human Perception and Performance, 1976, Vol. 2, 14-22.

19. Shiffrin, R.M., Pisoni, D.B. & Casteneda-Mendez, K. Is Attention Shared Between the Ears? Cognitive Psychology, 1974, Vol. 6, 190-215.

20. Lynn, R. Attention, Arousal and the Orienting Response. Oxford: Pergamon Press, 1966.

21. Jennings, J.R. & Hall, S.W. Recall, Recognition, and Rate: Memory and the Heart. Psychophysiology, 1980, Vol. 17, 37-45.

22. Cohen, R.A. & Waters, W.F. Psychophysiological Correlates of Levels and Stages of Cognitive Processing. Neuropsychologia, 1985, Vol. 23, 243-256.

23. Bartley, S.H. Fatigue. Perception and Motor Skills. 1981, Vol. 53, 958.

24. Broadbent, D.E. Is a Fatigue Test Now Possible? Ergonomics, 1979, Vol. 22, 1277-1290.

25. Diagnostic and Statistical Manual of Mental Disorders: American Psychiatric Association, 1980 (3rd. Ed.).

26. Sandyk, R. Anergia and Depression as Early Features of Multiple Sclerosis. South African Medical Journal. 1982, Vol. 20, 732.

27. Hallpike, J.P. Clinical aspects of M.S. Multiple Sclerosis. Hallpike, P., Adams, C.W.M. & Tourtellotte, W.W. Chapman & Hall, New York, 1983.

28. Sokolov, E.N. Perception and the Conditioned Reflex. Oxford: Pergamon, Press, 1963.

29. Waters, W.F. & Wright, D.C. Maintenance and Habituation of the Phasic Orienting Response to Competing Stimuli in Selective Attention. In Kimmel, Van Olst and Orlebeke (eds): The Orienting Reflex in Humans, New York: Lawrence Erlbaum Publishers, 1979.

30. Groves, P.M. & Thompson, R.F. Habituation: A Dual-Process Theory, Psychological Review, 1970, Vol. 77, 419-450.

31. Lacey, B.C. & Lacy, J.I. Studies of Heart Rate and Other Bodily Processes in Sensorimotor Behavior. In P.A. Obrist, A.H. Black, J. Brener, & L.V. DiCara (Eds.), Cardiovascular psychophysiology: Current issues in response mechanisms, biofeedback and methodology. Chicago: Aldine-Atherton, 1974.

32. McGuigan, F.J. & Schoonover, R.A. The Psychophysiology of Thinking: Studies of Covert Processes. New York: Academic Press, 1973.

33. Jennings, J.R. Memory Thought and Bodily Response. In Coles, M.G.H., Danchin, E. & Porges, S.W. (Eds.) Psychophysiology. Coles, Donchin, Porges. New York, Guilford Press, New York, 1986.

34. Jennings, J.R., Lawrence, B.E. & Kasper, P. Changes in Alertness and Processing Capacity in a Serial Learning Task. Memory and Cognition. 1978, Vol. 6, 45-63.

35. Graham, F.K. Distinguishing Among Orienting, Defense, and Startle Reflexes. In H.D. Kimmel, E.H. Van Olst, & J.F. Orlebeke (Eds). The Orienting Reflex in Humans. Hillsdale, N.J.,: Erlbaum, 1979.

36. Obrist, P.A., Howard, J.L., Lawler, J.E., Galosy, R.A., Meyers, K.A. & Gaebelein, C.J. The Cardiac-Somatic Interaction. In P.A. Obrist, A.H. Black, J. Brener, & L.V. DiCara (Eds), Cardiovascular psychophysiology: Current issues in response mechanisms, biofeedback and methodology. Chicago: Aldine-Atherton, 1974.

37. Kahneman, D., Tursky, B., Shapiro, D. & Crider, A. Pupillary, Heart Rate and Skin Resistance Changes During a Mental Task. Journal of Experimental Psychology, 1969, Vol. 79, 164-167.

38. Jennings, J.R. Information Processing and Concomitant Heart Rate Changes in the Overweight and Underweight. Physiological Psychology, 1975, Vol. 3, 290-296.

39. Silverstein, L.D. & Berg, W.K. Repetition and Distribution Effects on Memory: A psychophysiological analysis. Paper presented at the annual meeting of the Society for Psychophysiological Research, October 1977.

40. Tursky, B., Schwartz, G.E. & Crider, A. Differential Patterns of Heart Rate and Skin Resistance During a Digit-transformation Task. Journal of Experimental Psychology, 1970, Vol. 83, 451-457.

41. Kahneman, D. & Beatty, J. Pupil Diameter and Load on Memory. Science, 1966, Vol. 154, 1583-1585.

42. Siddle, D.A.T., Kyriacou, C., Heron, P.A. & Matthews, W.A. Effects of Changes in Verbal Stimuli on the Skin Conductance Response Component of the Orienting Response. Psychophysiology, 1979, Vol. 16, 34-40.

43. Yuille, J.L. & Harre, R.D. A Psychophysiological Investigation of Short-term Memory, Psychophysiology, 1980, Vol. 17, 432-430.

44. Cacioppo, J.T. & Petty, R.E. Electromyographic Specificity During Covert Information Processing. Psychophysiology, 1981, Vol. 18, 518-523.

45. Locke, J.L. & Fehr, F.S. Young Children's Use of the Speech Code in a Recall Task. Journal of Experimental Child Psychology, 1970, Vol. 10, 367-373.

46. Cole, R.A. & Young, M. Effect of Subvocalization on Memory for Speech Sounds. Journal of Experimental Psychology, 1975, Vol. 1, 772-774.

47. Craik, F.I.M. & Lockhart, F.S. Levels of Processing: a Framework for Memory Research. J. Verb. Learn. Verb. Behar. 1972, Vol. 11, 671-684.

48. Lezak, M.D. Neuropsychological Assessment. New York, Oxford University Press, 1983.

49. Heilman, K.M. & Valenstein, E. (eds). Clinical Human Neuropsychology, New York, Oxford University Press, 1985.

50. Luria, A.R. Higher Cortical Functions in Man, Basic Books, 1980, New York.

51. Barkley, R.A. Assessment of Attention Deficit Disorders in Children. Behavioral Assessment, 1987, Vol. 9, 207-234.

52. Barkley, R.A. Attention; Tramonthana, M., and Hooper, S. (Editors), Assessment Issues in Child Neuropsychology: New York, Plenum, 1988, 115-154.

53. Cohen, R. A., and Fisher, M. Neuropsychological Correlates of Fatigue Associated with Multiple Sclerosis, Journal of Clinical and Experimental Neuropsychology, 1988, Vol. 10(1), 48.

54. Cohen, R.A., Fennell E. & Bauer, R. Neuropsychological Correlates of Unipolar and Bipolar Affective Disorders, Paper presented at the International Neuropsychological Society, Houston, 1984.

55. Flor-Henry, P. & Yeudall, L.T. Neuropsychological Investigation of Schizophrenia and Manic-depressive Psychoses. In J. Gruzelier & P. Flor-Henry (eds), Hemisphere asymmetries of function in psychopathology. Amsterdam: Elsevier/North-Holland, 1979.

56. Taylor, M.A., Greenspan, B. & Abrams, R. Lateralized Neuropsychological Dysfunction in Affective Disorder and Schizophrenia. American Journal of Psychiatry, 1979, Vol. 136, 1031-1034.

57. Kleinsmith, L.J. & Kaplan, S. Paired Associative Learning as a Function of Arousal and Interpreted Interval. Journal of Experimental Psychology, 1963, Vol. 65, 190-194.

58. Folkard, S. Time of Day and Level of Processing. Memory and Cognition, 1979, Vol. 7, 247-252.

59. Folkard, S. & Greeman, A.L. Salience Induced Muscle Tension, and the Ability to Ignore Irrelevant Information. Quarterly Journal of Experimental Psychology, 1974, Vol. 26, 360-367.

60. Cohen, R.A. & Albers, H.E. Neuropsychological and Circadian Deficits Associated with Pituitary Craniopharyngioma, International Neuropsychological Society. Houston, February 1984.

Table 1. Important variables affecting attention

- Perceptual complexity
- Response demands
- Required cognitive operations
- Memory requirements
- Task length (Vigilance demands)
- Task or information salience
- Individual capacity differences
- Intrinsic factors affecting arousal

Table 2. Summary: mean differences derived from Duncan's Multiple Range Tests*

HR	Cue Covert processing Verbalization	**PL HSL**>LSL **HSL LSL**>PL HSL>LSL>PL
SC	Cue Covert processing Verbalization	PL>**HSL LSL** **HSL LSL**>PL HSL>LSL>PL
ST	Cue Covert processing Verbalization	**PL HSL LSL** **PL HSL LSL** HSL>**LSL PL**
EMG	Verbalization>covert processing>cue (**HSL LSL PL**)	

*Variables in bold type are not significantly different (=0.05).
PL = Phonemic level LSL = Low semantic level HSL = High semantic level

Table 3. Factors to be controlled in attentional assessment

- Multivariate assessment framework
- Range of tasks from different cognitive systems
- Serial / Multi-frame design
- Tasks varying in perceptual requirements
- Tasks varying in reponse production demands
- Means of quantifying demands and required effort
- Quantification of internal capacity limitations
- Correlation of physiological factors with task demands

FLOWCHART ILLUSTRATING SEQUENCE ON EACH TRIAL

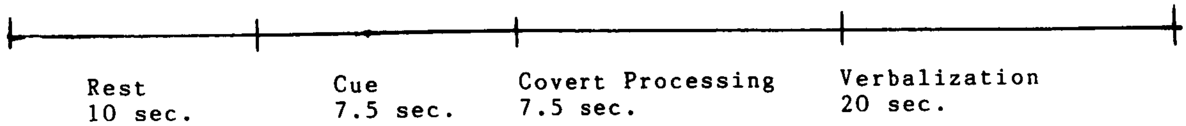


FIGURE 1. Flowchart illustrating stages for each trial of levels of processing paradigm

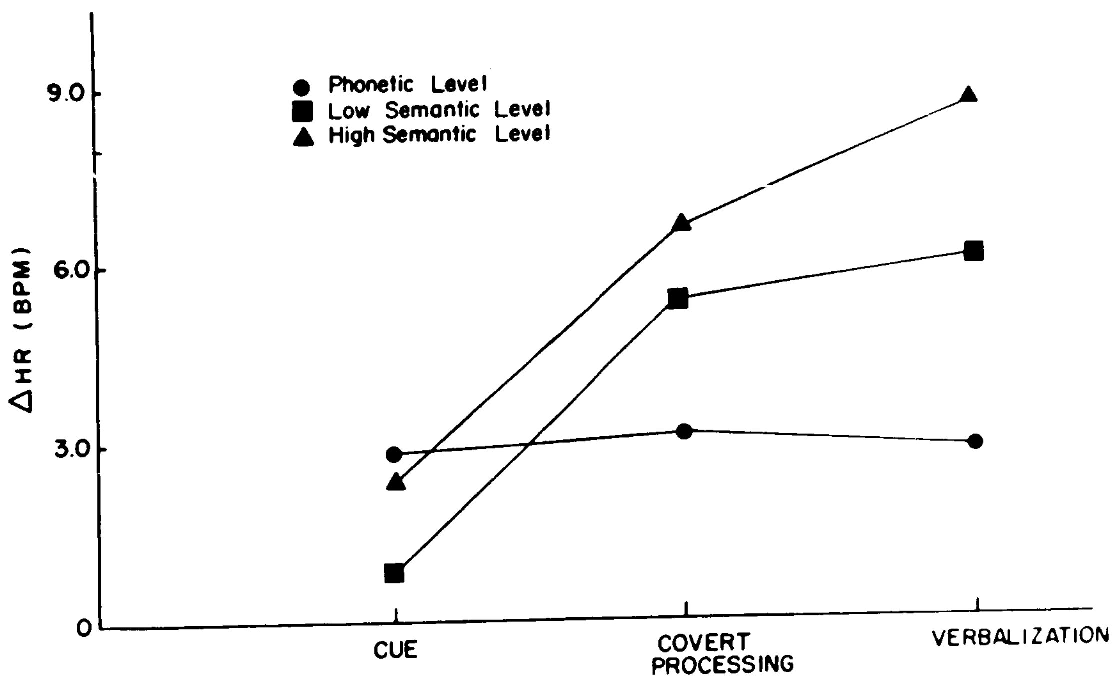


FIGURE 2. Heart rate change across the three stages of the task for each processing level

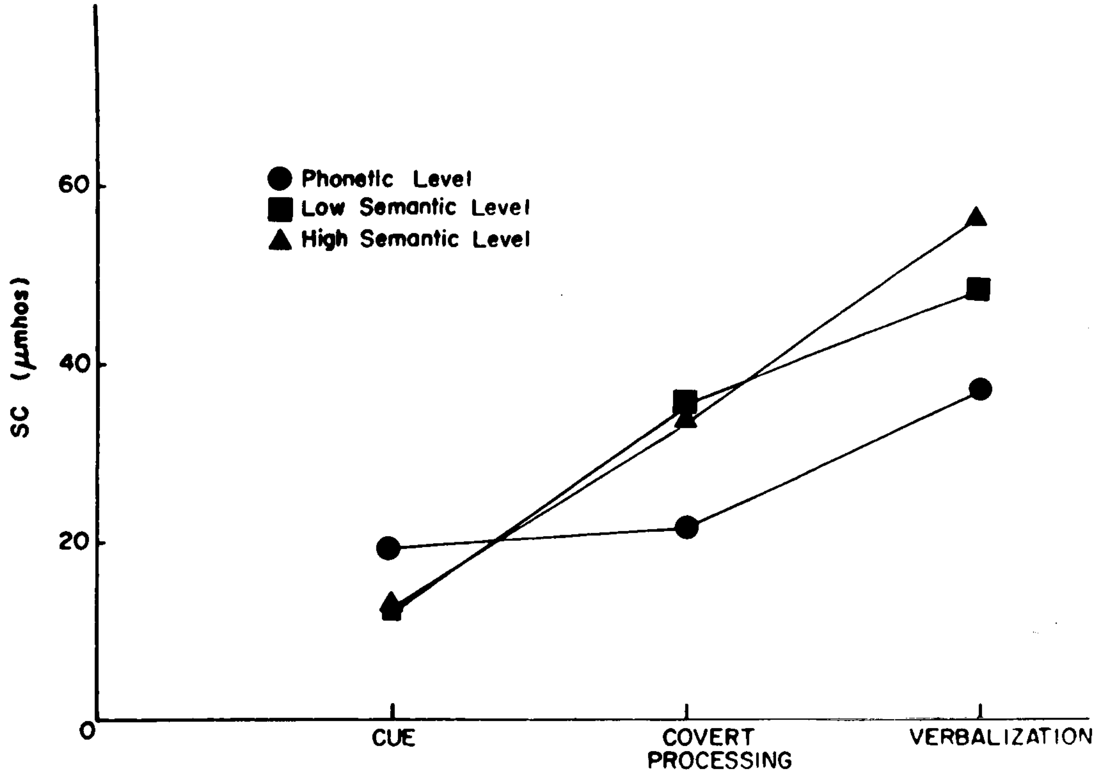


FIGURE 3. Skin conductance response across the three stages of the task for each processing level

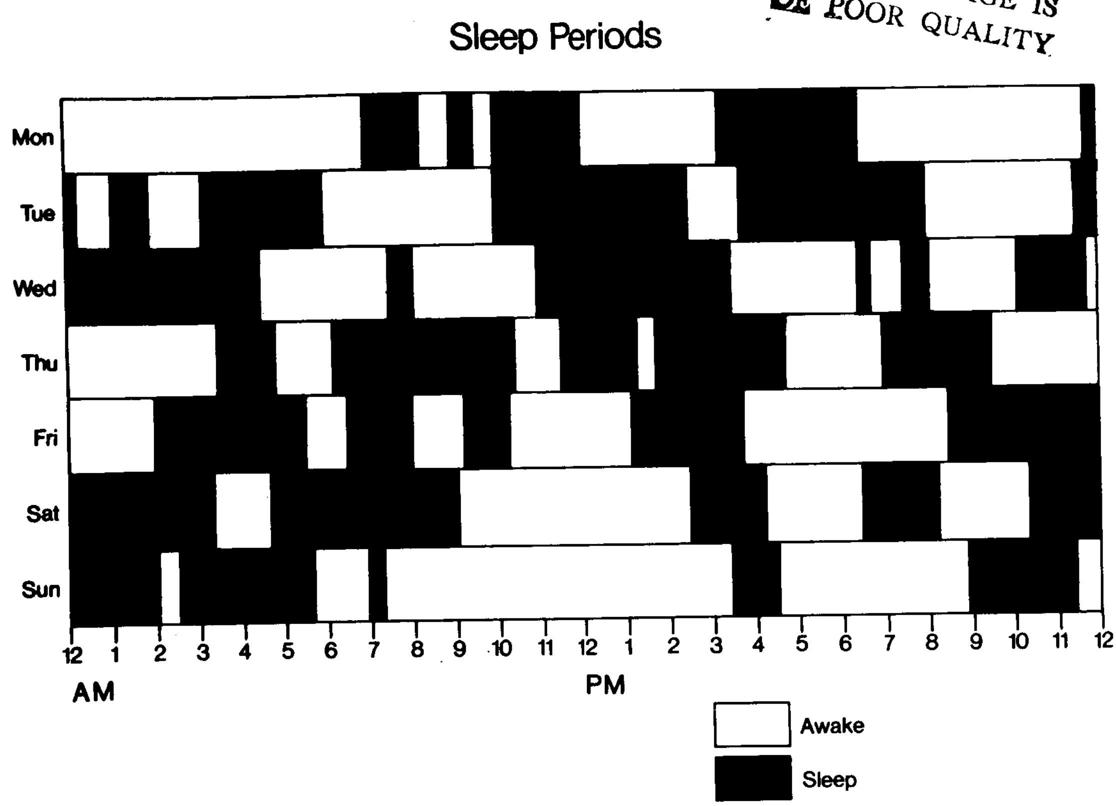


FIGURE 4. Sleep / Wake pattern for A.H., a patient with a history of craniopharyngioma

SUMMARY OF NEUROPSYCHOLOGICAL RESULTS

	FIRST EVALUATION 1ST TRIAL	FIRST EVALUATION 2ND TRIAL	SECOND EVALUATION
FSIQ	90		86
VERBAL IQ	95		94
PERFORMANCE IQ	86		81
INFORMATION	7		7
DIGIT SPAN	13		11
VOCABULARY	10		8
ARITHMETIC	11		10
COMPREHENSION	7		8
SIMILARITIES	6		8
PICTURE COMPLETION	6		6
PICTURE ARRANGEMENT	8		7
BLOCK DESIGN	10		9
OBJECT ASSEMBLY	6		6
DIGIT SYMBOL	5		3
MQ=	74		70
NEW PAIRED ASSOCIATES	0/4		0/4
LOGICAL STORIES	2.0		4.0
RAVLT	6/15		7/15
RECOGNITION	13/15		12/15
BOSTON NAMING	51/85		45/85
WORD GENERATION			
(WORDS/CATEGORY)	5	10	3
TRAIL A (SEC)	52	33	65
TRAIL B (SEC)	124	75	208
STROOP			
WORD	75	86	84
COLOR	54	42	53
INTERFERANCE TRIAL	32	23	26

N88-23388 5/8-53
140806
188.

THE N2-P3 COMPLEX OF THE EVOKED POTENTIAL AND HUMAN PERFORMANCE

Brian F. O'Donnell and Ronald A. Cohen
Department of Neurology
University of Massachusetts Medical School
Worcester, MA

MK 170541

When sensory receptors are stimulated, a series of negative and positive deflections time-locked to stimulus onset may be evoked in the electroencephalogram (EEG). Since these potentials are evoked by sensory stimulation, they are called sensory-evoked potentials (EPs). Because of the small magnitude of the EP in relation to ongoing background noise, many stimulus trials must be averaged to obtain a stable EP.

EP waveforms are quantitatively characterized in terms of components. Unfortunately, there is no consensus in the field as to the formal definition of a component (ref. 1). For the paradigms discussed in this paper, components are identified with specific positive and negative deflections in the averaged EP. The deflections are labeled by their polarity and order of appearance. Polarity of a deflection is either positive or negative, denoted by the prefixes "P" or "N". N1, for example, would be the first major negative deflection observed after presentation of an auditory stimulus. N1 generally occurs about 100 milliseconds (ms) after stimulus onset, and is for this reason sometimes labeled N100. Labeling components by their polarity and latency after stimulus onset ("N100", "P300") is another frequently used convention in the EP literature.

EP components are functionally categorized into two types, exogenous and endogenous. Exogenous components of the EP are primarily responsive to properties of the stimulus, such as duration, intensity, and frequency. Typically, exogenous components have short latencies (less than 100 ms after stimulus onset). They usually originate from the primary sensory pathways and projection areas. The morphology and scalp distribution of exogenous components vary greatly between stimulus modalities, and are relatively little affected by task demands.

Endogenous components of the EP vary with psychological factors such as task relevance, expectancies, and task difficulty. EPs associated with endogenous components are frequently referred to as event-related potentials (ERPs). In this paper, the properties of a set of endogenous components, the P3 complex, will be discussed. The P3, or P300, component has received continued experimental attention since it was first reported by Sutton, Braren, Tueting, Zubin and John (refs. 2 and 3). The P3 is a long latency, endogenous component of the evoked potential which can be elicited by auditory, visual, or somatosensory stimuli. In a typical paradigm, the P3 is evoked when a subject attends to rare target tones among a train of more frequently presented non-target tones. P3 usually appears at a latency between 250 and 800 ms after stimulus onset. It is generally preceded by a negative deflection (N2) and followed by a deflection whose polarity varies with scalp topography, the "Slow Wave" (ref. 4). These endogenous

C-4

components are shown in Figure 1. While N2 and P3 usually appear sequentially, they are dissociable. The topography of N2 is modality specific; that is, its peak amplitude appears at different locations on the scalp depending on modality of stimulation (refs. 5 and 6). P3 shows a modality non-specific scalp topography, with peak amplitude over the parietal area of the scalp. N2 appears to a stimulus mismatch whether or not the stimulus is task relevant, whereas the P3 response is attenuated or absent under these conditions (ref. 7). The neural generators of P3 are not known with any specificity. Evidence from depth electrode recordings and correlations with magnetic fields suggest that medial temporal lobe and frontal lobe structures may be involved (refs. 8 to 10).

This paper will address the responsivity of the N2 and P3 components of the EP (the N2-P3 complex) to factors modulating human performance. The first section reviews experimental factors and paradigms. The second and third sections examine the effects of brain dysfunction and pharmacological manipulations on the N2-P3 complex. The functional significance of the N2-P3 complex and its utility as a tool for probing human performance will then be discussed.

Factors Which Influence the N2-P3 Complex

Probability and Task Relevance

Variations in stimulus probability are associated with changes in N2-P3 amplitude (ref. 2). The effect of probability on P3 amplitude is enhanced when the stimuli are task relevant (ref. 11). When a stimulus is ignored, the P3 deflection that occurs (P3a) may represent a different component from the P3 deflection to a task-relevant stimulus (P3b) (ref. 4). A large P3 may be evoked without task demands when a rare tone is very disparate in intensity and frequency from a frequent tone (ref. 12). Task relevant stimuli are usually associated with N2-P3 activity even when the stimuli are equiprobable in relation to the irrelevant stimuli (ref. 2). N2 amplitude is less sensitive to task demands, suggesting that it may represent an automatic match-mismatch detection process (ref. 7). The amplitude of P3 is inversely related to stimulus probability, approximating its information content as defined by classical information theory ($-\log 2p$) (ref. 13). P3 amplitude to a feedback signal regarding a previous judgment on a target detection task is related to the joint probability of the initial stimulus and the subject's response, termed outcome probability (ref. 14) or contingent probability (ref. 13).

Sequential stimulus structure also contributes to N2-P3 amplitude. The first stimulus of a series elicits a N2-P3 complex. A tone preceded by one or more of the same tones shows diminished N2-P3 amplitude, and one preceded by a series of differing tones shows larger amplitude responses (ref. 15). K. Squires et al. (ref. 15) used a linear additive model defining expectancy as a combination of decaying memory for events, structure sequence, and global probability for up to fifth order stimulus sequences. The model accounted for 78% of the variance of N2-P3 amplitude. Duncan-Johnson and Donchin (ref. 11) similarly found that global probability and sequential structure had independent effects on the P3 complex.

In summary, global stimulus probability and stimulus sequence are

important determinants of the amplitude of the N2-P3 complex. These effects interact with the task-relevance of the stimulus. Task relevant stimuli produce a N2-P3 complex, and the effect of probability is greatly enhanced when stimuli are task-relevant. The joint effects of task relevance and probability provide an example of the sensitivity of electrophysiological measures to aspects of information processing and attentional reactivity not readily apparent from traditional psychological paradigms.

Orienting response

Both N2 and P3 have been associated with the orienting response (refs. 7, 16 and 17). The orienting response is elicited by a variation in stimulus properties, presumably because of a mismatch between the previous representation of the stimulus and the physical properties of the current stimulus. The response is manifested by a range of autonomic, somatic, and EEG changes (ref. 17). The N2-P3 complex fits this model in its reactivity to stimulus change and probability. It diverges from the classical orientation response in its resistance to habituation, even over prolonged periods of time (refs. 18 and 19). One difficulty in making comparisons between the N2-P3 complex and the orienting response is that few studies have used both autonomic and EP measures simultaneously in classical orienting paradigms. A second difficulty is that experiments designed to elicit the N2-P3 complex use short inter-stimulus intervals and task relevant stimuli, while the orienting response classically has not been associated with explicit task demands (ref. 20). A recent study by Rosler (ref. 21) compared N2, P3, skin conductance and HR to rare and frequent visual stimuli. The results indicated that these different response modalities were related to different aspects of task demands and stimulus properties. Rosler concluded the ensemble of autonomic and EP measures was not part of a single orienting reflex, but rather was sensitive to different stages of information processing. Late negative waves occurring after the N2-P3 complex (Slow Wave, "O" wave, CNV) have been argued to be more closely related to the orienting response (ref. 17 and 22).

N2-P3 and motor response

N2 latency, P3 latency, and reaction time (RT) tend to be correlated, particularly when accuracy of response is stressed over speed of response (ref. 23). The P3 component, however, occurs too late after stimulus onset to be concurrent with stimulus discrimination and a precondition for response selection and execution. Ritter and colleagues (ref. 24) have argued that N2 is a better time marker for stimulus discrimination. Goodin and colleagues (ref. 25), however, report data (using EMG onset as a measure of reaction time)which suggest that N2 may also be too late in time to directly index stimulus discrimination. It is possible that the processes represented by N2 and P3, as well as response selection, are initiated in parallel by early stimulus analysis, but the response selection is not necessarily contingent upon N2-P3 related activities in the nervous system. A recent experiment by Goodin and colleagues (ref. 26) demonstrated that P3 and an earlier endogenous component, P165 (Figure 1), were synchronized with both stimulus appearance and response onset as measured by EMG activity, while N2 was more synchronized with stimulus onset than response onset. These results provide further evidence that N2 may represent an independent process from P3, even though they appear sequentially in the averaged EP.

Stimulus evaluation and signal detection

Stimulus evaluation. Stimulus intensity is inversely related to P3 latency (ref. 27). Increased difficulty of discrimination is associated with increased N2 and P3 latency (refs. 28 to 32). Task demands which increase the complexity of stimulus evaluation increase P3 latency and RT, while task demands which increase the difficulty of response selection increase RT latency without affecting P3 latency (ref. 32). Variations in visual stimulus intensity, contrast, and complexity have additive effects on P3 latency (ref. 31). These results have led several investigators to propose that P3 latency provides an index of stimulus discrimination in the nervous system (refs. 23 and 29). Because RT is not temporally contingent on P3, however, it appears more likely that P3 latency represents further processing of a stimulus contingent on initial discrimination, and parallel to response selection.

Signal detection. The effects of observer sensitivity and decision confidence on P3 latency have been studied by a number of investigators. N1 has been related to quantity of signal information received by the subject, while P3 characteristics reflect decision confidence (ref. 33). P3 amplitude increases, and latency decreases, with increasing confidence for correct detection (Hit) of a signal (refs. 34 to 36). In general, false alarms, misses, and correct rejections in signal detection tasks are associated with smaller amplitude P3s. P3 responses will occur to confident false alarms (ref. 33). Correct rejections generate P3s only when signals are highly detectable and signal-absent trials are rare (ref. 35). When signals are of low detectability, probability of presentation has little effect on P3 amplitude (refs. 34 and 35). In a study of signal detection and recognition, P3 amplitude increased and latency decreased as a function of both signal detection and recognition, while N1 only varied with signal detection (ref. 36).

In summary, while P3 probably does not provide a direct marker for the time of stimulus discrimination in the nervous system, it does provide a sensitive measure of the process of stimulus evaluation. P3 latency increases with difficulty of a discrimination. P3 amplitude, on the other hand, reflects decision confidence related to both detection and recognition of signal. The N2-P3 complex in conjunction with RT provides a powerful paradigm for the chronometric analysis of stimulus processing, decision processes and response generation in the human central nervous system (CNS).

Mental load

The findings that the amplitude of P3 was modulated by task relevance and attentional focus, and signal its latency to stimulus evaluation led investigators to link P3 amplitude to the conscious deployment of limited capacity processing resources (refs. 37 and 38). Several lines of research are consistent with this formulation, and suggest that P3 is sensitive to the mental load presented by a task.

The Stroop interference effect, which appears to be due to response interference, prolongs RT without affecting P3 latency (ref. 39). Dual task performance diminishes P3 amplitude on the primary task when the secondary task makes demands on perceptual resources, though not when further demands

are placed on elaboration of a response. RT is responsive to both types of demands (refs. 40 and 41). Wickens and colleagues (ref. 38) hypothesized that if processing resources allocated to a primary and secondary task were reciprocal, this relationship should be reflected in variations in P3 amplitude to stimuli in both tasks. Using visual tracking as the primary task, and an auditory oddball sequence as the secondary task, they compared P3 amplitude to stimuli within each task. As the resource demands of the primary task were increased, P3 amplitude evoked by primary task events increased, whereas those elicited by the auditory stimuli used in the secondary task decreased. A distinction between the responsivity of N2 and P3 amplitude to task relevant and irrelevant workload was reported by Horst and colleagues (ref. 42). When subjects were required to monitor multiple visual readouts, increasing workload was associated with increased negativity in the N2 region of the waveform, regardless of whether the readout was currently task relevant. In the P3 regions of the EP, however, increased workload only affected component amplitude to attended, task-relevant stimuli.

Automatic and controlled processing in visual search tasks (ref. 43) have also been investigated using EP and RT paradigms. N2-P3 amplitude was comparable in automatic and controlled tasks in two studies, while both P3 and RT latencies were shortened in the automatic task (refs. 44 and 45). Memory set size did have an effect on amplitudes, however: N2 amplitude was smaller, and P3 amplitude larger, with increased memory set size (ref. 45). These results suggest that practicing a controlled mapping task (comparing a stimulus to a constant set of items in memory) may reduce the slope of stimulus evaluation and reaction time on memory set size to zero, but the task still requires perceptual resources for performance.

These initial studies suggest that P3 amplitude reflects the mental demands on limited-capacity perceptual resources. In conjunction with RT measures, it may provide a means of differentiating perceptual and response related resource demands involved in performance of specific tasks.

Learning and Memory

P3 amplitude is enhanced to stimuli which are examples of an infrequently occurring category in a series when the stimuli share no common physical properties (refs. 23 and 46). Such results suggest that learned categories in long term memory can be probed by N2-P3 responsivity. The learning process has been experimentally investigated by requiring a subject to learn, either intentionally or not, a set of items, and then measuring the magnitude and latency of P3 of items correctly recognized or missed on a subsequent exposure. P3s to recognized stimuli were larger in amplitude and shorter in latency than those to unrecognized stimuli or distractors, independent of relative probabilities. These results were interpreted to be consistent with the hypothesis that recognized items are more familiar, hence more discriminable, then unrecognized items (refs. 47 and 48). On repeated learning tests, P3 latency becomes shorter and P3 amplitude larger for correctly identified targets (ref. 48).

Several studies have examined whether P3 amplitude or latency to a stimulus on initial exposure predicts subsequent recognition performance. The hypothesis advanced by Donchin (ref. 49) that P3 reflects the process of context or schema updating suggests that stimuli associated with enhanced P3

activity should be more memorable than those that are not. Tests of this hypothesis have not led to consistent results. Sanquist et al. (ref. 47) reported an apparent (but statistically untested) increased amplitude during semantic processing of items which were later recognized. Fabiani, Karis and colleagues (refs. 50 and 51) reported a similar effect, but only when the subjects used a rote rehersal strategy, or no strategy at all, in the process of learning the material; elaborative strategies produced no P3 enhancement. P3 latency, but not amplitude, on repeated exposures of a list was shorter for words later recognized than to those that were not recognized. This effect may have been due to increased familiarity and discriminability of recognized words over repeated trials. In a continuous recognition task, P3 amplitude on initial exposure has been found to be predictive of later correct recognition (ref. 52). These results suggest that the latency or amplitude of P3 response may predict later recognition performance, although the nature and strength of this effect may be paradigm specific.

Brain Dysfunction and the N2-P3 Complex

The N2-P3 complex has been studied in relation to normal aging, in psychopathology, and in neurological brain disorders. The most intensively studied clinical populations include patients with dementing disorders, schizophrenia, and depression. Variations of oddball paradigms, without or without RT measures, have been the most frequently used EP tests. The P3 component has been the most generally measured EP component in these disorders, although some studies also report characteristics of other components.

Aging

After adolescence, N2 and P3 latency show a continuous increase in latency. The rate of prolongation is about 1 to 2 ms per year. A decrease in P3 amplitude has also been reported (refs. 53 and 54).

Dementia

Dementing disorders such as Alzheimer's disease, multi-infarct dementia, and Parkinson's disease are usually accompanied by prolongation of N2 and P3 (refs. 54 to 58).

Psychiatric disorders

Both N2 and P3 amplitudes have been consistently reported to be reduced in amplitude in schizophrenia (refs. 59 to 63) and depression (refs. 55 and 61). N2 and P3 latency are usually reported to be within normal limits in these disorders, although there have been reports of mild slowing in schizophrenic patients (refs. 55 and 63). Since N2 and P3 latency are usually within normal range in schizophrenia, while RT is slowed, this particular type of psychopathology may reflect disturbances of response selection and execution more than stimulus evaluation (ref. 64).

Correlation of N2-P3 with Neuropsychological Measures

Few EP studies provide behavioral or intellectual descriptions of patient groups beyond diagnosis. In the case of dementia, groups under study were often heterogeneous in diagnosis as well as severity. Specific intellectual or psychiatric disturbances relevant to such constructs as attention, learning, or degree of depression are seldom measured or correlated with specific EP changes. Consequently, the specific behavioral referents of variations in the N2-P3 complex due to brain dysfunction remain to be elucidated. Several recent studies of Parkinson's disease, a neurological disorder associated with varying degrees of motoric, intellectual, and psychiatric disturbance, have examined such patterns. The latency of P3 in Parkinson's disease is correlated with mental tests requiring cognitive effort and learning, and is less related to general measures of IQ, immediate memory span, depression or motor dysfunction (refs. 57 and 58). These results suggest that N2 and P3 changes associated with brain dysfunction may index specific types of cognitive and behavioral disturbance, in the same way that N2 and P3 characteristics in experimental paradigms vary with specific types of task demands.

Summary

The N2-P3 complex is delayed over the course of normal aging, and further delayed in dementing disorders associated with diffuse brain damage. In Parkinson's disease, P3 latency changes correlate with deficits in learning and tasks requiring cognitive effort. Psychiatric disorders, on the other hand, are consistently associated with reduction in N2-P3 amplitude, with relatively normal component latencies. This pattern of results may indicate that N2-P3 latency prolongation is a marker for clinically significant slowing of mental processes, or memory deficits, while diminished amplitude is associated with disorders affecting attention, motivation or arousal. The finding that seizure patients show increased P3 amplitude is consistent with the notion that P3 amplitude is a measure of CNS arousal (ref. 65).

Pharmacological Effects

The N2-P3 complex is differentially reactive to CNS stimulants and anticholinergic agents. Methylphenidate speeds RT without affecting P3 latency in young adults and children with attention disorders. This pattern suggests that methylphenidate speeds response generation, but does not affect stimulus evaluation processes (ref. 66). D-amphetamine, on the other hand, reduces both P3 latency and RT latency. These effects were not reduced by administering propranalol (ref. 67). The effect of d-amphetamine on P3 latency did not interact with stimulus complexity.

Scopolamine, an anti-cholinergic agent, slows both P3 and RT latency (ref. 66). At high levels, scopolamine abolishes P3 response and causes severe learning deficits, despite accurate task performance and retained immediate memory span (ref. 68).

These results again demonstrate the power the N2-P3, in conjunction with reaction time, to provide chronometric probes of the locus of variation in human performance. The effects of anti-cholinergic agents on the N2-P3

complex suggest that N2-P3 slowing may reflect breakdown in attentional and learning processes, similar to its significance in clinical disorders of the CNS.

The Cognitive Significance of the N2-P3 Complex

The P3 component has been described as indexing uncertainty (ref. 2), significance, information delivery (ref. 3), orienting (ref. 16) expectancy (ref. 15), equivocation (ref. 69), stimulus evaluation (refs. 23 and 29), context or schema updating (ref. 49), and value or meaning (ref. 1). This multiplicity of hypotheses regarding the functional significance of P3 reflects the diverse range of experimental manipulations which can affect P3 amplitude, latency, or both features. As is evident from the preceding review, the N2 component is reactive to many of the same factors as P3, although it may represent a more automatic phase of stimulus evaluation. Donchin (ref. 49) suggested that the P3 component may represent the CNS equivalent of a subroutine, which is invoked in a variety of cognitive operations. Alternatively, since the P3 may not consist of a single component, but rather the sum of a number of components overlapping in time (ref. 1), the characteristics of the P3 complex may index more than a single CNS function.

A model of P3 amplitude which assumes multiple determinants has been developed by Johnson (1986). Johnson (ref. 70) proposed that P3 amplitude is determined by three factors: subjective probability, stimulus meaning, and information transmission. Subjective probability is a joint function of global and sequential expectancies, as previously modeled by K. C. Squires and colleagues (ref. 15). Stimulus meaning is a function of task complexity, stimulus complexity, and stimulus value. Johnson proposed that subjective probability and stimulus meaning have an additive relationship, while both have a multiplicative relationship with information transmission. He makes the intriguing suggestion that subjective probability is an automatic process, while stimulus meaning is a controlled process.

The Assessment of Human Performance

The utility of the N2-P3 complex as a probe of CNS processes associated with stimulus evaluation, attentional variation, and mental load has been repeatedly demonstrated over the past two decades. Clinical and pharmacological evidence suggests that these measures are also sensitive to global changes in the information processing capacity of the CNS due to brain dysfunction. The effect of common stressors on human performance, such as fatigue, boredom, noise, or sleep deprivation on the N2-P3 complex has received much less attention. Further research is needed to elucidate how such stressors impact on the N2-P3 complex, and how this impact influences task performance. The inclusion of subjective measures of mood, arousal, and personality as setting variables in experiments may permit the development of multifactorial models of the determinants of psychophysiological response. Unlike machine information processing systems, human performance is modulated by biological and personality factors. Psychophysiological measures may provide markers for such influences.

In the evaluation of human performance, behavioral and subjective measures of performance are readily available. As Donchin (ref. 71) has argued, given the constraints and costs imposed by EP assessment of CNS function, EPs should be used only when they provide information which is not easily available from traditional indices of performance. The foregoing review of the N2-P3 complex suggests several applications in which unique information can be derived from EP measurement.

1. Evaluation of the time course of stimulus evaluation processes as distinct from response selection and execution.

2. Electrophysiological assessment of the attentional impact of infrequent events.

3. Measurement of workload specifically related to perceptual capacity. The auditory oddball task provides a relatively unobtrusive measure of secondary task processing. In addition, P3 amplitude may provide a direct measure of perceptual workload.

4. Characterizing the salience of events to an operator without requiring a behavioral response.

5. Identifying the time points in sensory and perceptual processing when pharmacological manipulations become effective.

6. Assessing the integrity of brain function.

Methodological Considerations

EP component identification and analysis

A variety of analytic techniques have been used to identify and measure components of the N2-P3 complex. The lack of consensus on identification and quantitative characterization of EP components, and the difficulty of discriminating variations in the latency of these components from single trials, has been a cause of continued concern and the application of diverse analytic techniques to EPs. (See Sutton and Ruchkin, 1984 (ref. 1), for an excellent discussion of the problems of component definition.) Popular analytic approaches include Woody filtering, subtraction waveforms, digital filtering, principal components analysis, peak-picking, and single-trial latency adjustment. Despite the obvious methodological concerns demonstrated by investigators, however, the experimental and clinical effects reviewed above are remarkably robust.

The most serious problem in reviewing and integrating studies in the literature is not, in the opinion of this reviewer, the difficulty in identifying the central phenomena of interest (although mapping the N2-P3 complex onto experimental factors and mental functions remains a vigorous and productive enterprise after two decades of activity). Rather, it is the tendency of experimenters to focus a priori on components of interest, and ignore other potentially informative components in the EP waveforms. Consequently, it is not unusual to read studies involving similar experimental manipulations which focus on P3 measures, and ignore earlier components, or conversely, measure early components, such as processing

negativities, without measuring later components. As a basic guideline, given the differential reactivity of EP components to stimulus properties and task demands, the major components of the N2-P3 complex (N2, P3, Slow Wave) should be measured, as well as at least one representative exogenous component (e.g. P100 with visual stimuli; N1 with auditory stimuli). Averaged EPs for each experimental conditions should be displayed before transformations such as principal components analysis are used. Indices of behavioral performance should be used in conjunction with EP responses when variations in task demands occur that may impact on response selection. A survey of papers presented at the Eighth International Conference on Event-Related Potentials of the Brain (ref. 72) suggests the field is moving toward greater specificity of measurement applied over the entire recorded EP epoch.

Ecologically Valid Experiments

The first phase of N2-P3 investigations, extending from perhaps 1965 to 1980, generally used stimuli with simple physical properties (e.g. tones, clicks, simple figures) and varied the stimuli on precise dimensions (e.g. intensity, probability, frequency). The benefit of this approach was a high degree of replicability across different laboratories, and the easy application of psychophysical, signal detection, and information processing paradigms. Moreover, since information processing was the dominant model of interpretation, semantic qualities of stimuli were not easily incorporated into analysis. Since the late 1970s, however, increasingly complex linguistic and visual stimulus paradigms have been utilized, presumably as consequence of investigators' confidence in their understanding of the basic characteristics of the N2-P3 complex. As the functional characteristics of these components have become understood, they have begun to be used as a tool for the understanding of mental processes, rather than being the explicit object of inquiry in an experiment. The evolution of EPs from an object of inquiry to a tool of inquiry has important implications for the investigation of human performance. Until this evolution occurred, application of EP measures to task analysis in engineering psychology would be a uninterpretable.

Developing more naturalistic tasks and environments will be an important step in using EPs to probe the CNS mechanisms modulating human performance. The constraints of EP analysis (the use of electrodes, electrical shielding, physiological amplifiers, analog or digital recording), the need for many trials to accrue an interpretable average, and the short time window of investigation limit the applicability of this technique. When the technique can be applied to a task, the stimuli, temporal frame, and environmental context should be as close as possible to the performance environment of interest.

Prediction of performance

The N2-P3 complex has usually been correlated with behavioral measures recorded concurrently in time. Prediction of subsequent human performance levels has seldom been a focus of investigation. It would be of great interest if properties of the N2-P3 complex might reflect an individual's general attentional or cognitive capabilities, and whether alterations in the N2-P3 complex in a serial task might reflect the probability of

subsequent lapses in attention. The sensitivity of the N2-P3 complex to brain dysfunction in clinical populations suggests it might show a similar sensitivity to diffuse changes in the CNS system in healthy individuals under unusual stress.

Summary

Two decades of productive research have demonstrated that the N2-P3 complex, and other endogenous components of the human EP (ref. 73), provide a set of tools for the investigation of human perceptual and cognitive processes. These multidimensional measures of CNS bioelectrical activity respond to a variety of environmental and internal factors which have been experimentally characterized. Their application to the analysis of human performance in naturalistic task environments is just beginning. Converging evidence suggests that the N2-P3 complex reflects processes of stimulus evaluation, perceptual resource allocation, and decision-making that proceed in parallel, rather than in series, with response generation.

Utilization of these EP components may provide insights into the CNS mechanisms modulating task performance unavailable from behavioral measures alone. The sensitivity of the N2-P3 complex to neuropathology, psychopathology, and pharmacological manipulation suggests that these components might provide sensitive markers for the effects of environmental stressors on the human CNS.

Acknowledgements

We would like to thank the following institutions for support: NIA Grant 1-P50-0AG05134, Alzheimer Disease Research Center; the Friedman Foundation; the University of Massachusetts Medical Center Scientific Council; and the Sterling Morton Charitable Trust.

References

1. Sutton, S., Ruchkin, D.S.: The Late Positive Complex: Advances and New Problems. In Karrer, R., Cohen, J., Tueting, P. (Eds.), Brain and Information: Event-Related Potentials. Annals of the New York Academy of Sciences, vol. 425, 1984, pp. 1-23.

2. Sutton, S., Braren, R., Zubin, J., John, E.R.: Evoked-Potential Correlates of Stimulus Uncertainty. Science, vol. 150, 1965, pp. 1187-1188.

3. Sutton, S., Tueting, P., Zubin, J., John, E.R.: Information Delivery and the Sensory Evoked Potential. Science, vol. 155, 1967, pp. 1436-1439.

4. Squires, N.K., Squires, K.C., Hillyard, S.A.: Two Varieties of Long-Latency Positive Waves Evoked by Unpredictable Auditory Stimuli in Man. Electroencephalography and Clinical Neurophysiology, vol. 38, 1975, pp. 387-401.

5. Simson, R., Vaughan, H.G., Ritter, W.: The Scalp Topography of Potentials Associated with Missing Visual or Auditory Stimuli. Electroencephalography and Clinical Neurophysiology, vol. 40, 1976, pp. 33-42.

6. Simson, R., Vaughan, H.G., Ritter, W.: The Scalp Topography of Potentials in Auditory and Visual Discrimination Tasks. Electroencephalography and Clinical Neurophysiology, vol. 42, 1977, pp. 528-535.

7. Naatanen, R., Simpson, M., Loveless, N.E.: Stimulus Deviance and Event-Related Brain Potentials. Biological Psychology, vol. 14, 1982, pp. 53-98.

8. Squires, N.K., Halgren, E., Wilson, C. and Crandall, P.: Human Endogenous Limbic Potentials: Cross-modality and Depth/Surface Comparisons in Epileptic Subjects. In Gaillard, A.W.K., Ritter, W. (Eds.), Tutorials in Event Related Potential Research: Endogenous Components. New York: North-Holland, 1983, pp. 217-232.

9. Okada, Y.C., Kaufman, L., Williamson, S.J.: The Hippocampal Formation as a Source of the Slow Endogenous Potentials. Electroencephalography and Clinical Neurophysiology, vol. 55, 1983, pp. 417-426.

10. Wood, C.C., McCarthy, G.: A Possible Frontal Lobe Contribution to Scalp P300. In Rohrbaugh, J.W., Johnson, R., Parasuraman, R. (Eds.), Eighth International Conference on Event-Related Potentials of the Brain: Research Reports. Stanford, 1986, pp. 164-166.

11. Duncan-Johnson, C., Donchin, E.: On Quantifying Surprise: The Variation of Event-Related Potentials with Subjective Probability. Psychophysiology, vol. 14, 1977, pp. 456-467.

12. Squires, N.K., Sanders, D., Wanser, R.: Comparison of Attend and Non-

Attend Paradigms for the Evaluation of ERP Changes in Normal Aging and Neurological Dysfunction. In Rohrbaugh, J.W., Johnson, R., Parasuraman, R. (Eds.), Eighth International Conference on Event-Related Potentials of the Brain: Research Reports. Stanford, 1986, pp. 137-139.

13. Campbell, K.B., Courchesne, E., Picton, T.W., Squires, K.C.: Evoked Potential Correlates of Human Information Processing. Biological Psychology, vol. 8, 1979, pp. 45-68.

14. Friedman, D. B., Hakarem, G., Sutton, S., Fleiss, J. L.: Effect of Stimulus Uncertainty on the Pupillary Dilatation Response and the Vertex Evoked Potential. Electroencephalography and Clinical Neurophysiology, vol. 34, 1973, pp. 475-484.

15. Squires, K.C., Wickens, C., Squires, N.K., Donchin, E.: The Effect of Stimulus Sequence on the Waveform of the Cortical Event-Related Potential. Science, vol. 193, 1976, pp. 1142-1146.

16. Ritter, W., Vaughan, H.G., Costa, L.D.: Orienting and Habituation to Auditory Stimuli: A Study of Short Term Changes in Averaged Evoked Responses. Electroencephalography and Clinical Neurophysiology, vol. 25, 1968, pp. 550-556.

17. Rohrbaugh, J.W. The Orienting Reflex: Performance and CNS Manifestations. In R. Parasuraman (Ed.), Varieties of Attention. New York: Academic Press, 1984.

18. Pritchard, W.S., Brandt, M.E., Shappell, S.A., O'Dell, T., Barratt, E.S.: No Decrement in Visual P300 Amplitude During Extended Performance of the Oddball Task. International Journal of Neuroscience, vol. 29, 1986, pp. 199-204.

19. Courchesne, E., Courchesne, R.Y., Hillyard, S.A.: The Effect of Stimulus Deviation on P3 Waves to Easily Recognized Stimuli. Neuropsychologia, vol. 16, 1978, pp. 189-199.

20. Donchin, E., Heftley, E., Hillyard, S.A., Loveless, N., Maltzman, I., Ohman, A., Fosler, F., Ruchkin, D., Siddle, D.: Cognition and Event-Related Potentials: II. The Orienting Reflex and P300. In Karrer, R., Cohen, J., Tueting, P. (Eds.), Brain and Information: Event Related Potentials. Annals of the New York Academy of Sciences, vol. 425, 1984, pp. 39-57.

21. Rosler, F.: Central and Peripheral Correlates of Orienting and Habituation. In Rohrbaugh, J.W., Johnson, R., Parasuraman, R. (Eds.), Eighth International Conference on Event-Related Potentials of the Brain: Research Reports. Stanford, 1986, pp. 254-256.

22. Loveless, N.: Event-Related Slow Potentials of the Brain as Expressions of the Orienting Function. In H. D. Kimmel, E. H. van Olst, J. F. Orlebeke, (Eds.), The Orienting Reflex in Humans. New York: Erlbaum, 1979.

23. Kutas, M., McCarthy, G., Donchin, E.: Augmenting Mental Chronometry:

The P300 as a Measure of Stimulus Evaluation Time. Science, vol. 197, 1977, pp. 792-795.

24. Ritter, W., Vaughan, H.G., Friedman, D.: A Brain Event Related to the Making of a Sensory Discrimination. Science, vol. 203, 1979, pp. 1358-1361.

25. Goodin, D.S., Aminoff, M.J.: The Relationship Between the Evoked Potential and Brain Events in Sensory Discrimination and Motor Response. Brain, vol. 107, 1984, pp. 241-251.

26. Goodin, D.S., Aminoff, M.J., Mantle, M.M.: Subclasses of Event-Related Potentials: Response-Locked and Stimulus-Locked Components. Annals of Neurology, vol. 20, 1986, pp. 603-609.

27. Papanicolaou, A.C., Loring, D.W., Raz, N., Eisenberg, H.M.: Relationship Between Stimulus Intensity and the P300. Psychophysiology, vol. 22, 1985, pp. 326-329.

28. Ritter, W., Simson, R., Vaughan, H.G.: Association Cortex Potentials and Reaction Time in Auditory Discrimination. Electroencephalography and Clinical Neurophysiology, vol. 33, 1972, pp. 547-555.

29. Squires, N.K., Donchin, E., Squires, K.C.: Bisensory Stimulation: Inferring Decision-Related Processes from the P300 Component. Journal of Experimental Psychology: Human Perception and Performance, vol. 3, 1977, pp. 299-315.

30. Goodin, D.S., Squires, K.C., Starr, A.: Variations in Early and Late Event-Related Components of the Auditory Evoked Potential with Task Difficulty. Electroencephalography and Clinical Neurophysiology, vol. 55, 1983, pp. 680-686.

31. Walton, P., Halliday, R., Naylor, H., Callaway, E.: Stimulus Intensity, Contrast and Complexity Have Additive Effects on P3 Latency. In Rohrbaugh, J.W., Johnson, R., Parasuraman, R. (Eds.), Eighth International Conference on Event-Related Potentials of the Brain: Research Reports. Stanford, 1986, pp. 409-411.

32. McCarthy, G., Donchin, E.: A Metric for Thought: A Comparison of P300 Latency and Reaction Time. Science, vol. 211, 1981, pp. 77-80.

33. Squires, K.C., Hillyard, S.A., Lindsay, P.A.: Vertex Potentials Evoked During Auditory Signal Detection: Relation to Decision Criteria. Perception & Psychophysics, vol. 14, 1973, pp. 265-272.

34. Squires, K.C., Squires, N.K.: Vertex Evoked Potentials in a Rating-Scale Detection Task: Relation to Signal Probability. Behavioral Biology, vol. 13, 1975, pp. 21-34.

35. Squires, K.C., Squires, N.K., Hillyard, S.A.: Decision-Related Cortical Potentials During an Auditory Signal Detection Task with Cued Observation Intervals. Journal of Experimental Psychology: Human Perception and Performance, vol. 1, 1975, pp. 268-279.

36. Parasuraman, R., Beatty, J.: Brain Events Underlying Detection and Recognition of Weak Sensory Signals. Science, vol. 210, 1980, pp. 80-83.

37. Posner, M.I.: Psychobiology of Attention. In Gazzaniga, M.S., Blakemore, C. (Eds.), Handbook of Psychobiology. New York: Academic Press, 1975.

38. Wickens, C., Kramer, A., Vanasse, L., Donchin, E.: Performance of Concurrent Tasks: A Psychophysiological Analysis of the Reciprocity of Information-Processing Resources. Science, vol. 226, 1983, pp. 1080-1082.

39. Warren, L.R., Marsh, G.R.: Changes in Event Related Potentials During Processing of Stroop Stimuli. International Journal of Neurosciences, vol. 9, 1979, pp. 217-223.

40. Israel, J.B., Chesney, G.L., Wickens, C.D., Donchin, E.: P300 and Tracking Difficulty: Evidence for Multiple Resources in Dual-Task Performance. Psychophysiology, vol. 17, 1980, pp. 259-273.

41. Israel, J.B., Wickens, C.D., Chesney, G.L., Donchin, E.: The Event-Related Brain Potential as an Index of Display-Monitoring Workload. Human Factors, vol. 22, 1980, pp. 211-224.

42. Horst, R.L., Munson, R.C., Ruchkin, D.S.: ERP Processing Negativities Related to Workload. In Rohrbaugh, J.W., Johnson, R., Parasuraman, R. (Eds.), Eighth International Conference on Event-Related Potentials of the Brain: Research Reports. Stanford, 1986, pp. 350-352.

43. Schneider, W., Schiffrin, R. M.: Controlled and Automatic Human Information Processing: I. Detection, Search and Attention. Psychological Review, vol. 84, 1977, pp. 1-66.

44. Hoffman, J.E., Simons, R.F., Houck, M.R.: Event-Related Potentials During Controlled and Automatic Target Detection. Psychophysiology, vol. 20, 1983, pp. 625-632.

45. Kramer, A., Schneider, W., Fisk, A., Donchin, E.: The Effects of Practice and Task Structure on the Components of the Event-Related Brain Potential. Psychophysiology, vol. 23, 1986, pp. 33-47.

46. Neville, H.J., Snyder, E., Woods, D.L., Galambos, R.: Recognition and Surprise Alter the Human Visual Evoked Response. Proceedings of the National Academy of Sciences USA, vol. 79, 1982, pp. 2121-2123.

47. Sanquist, T.F., Rohrbaugh, J.W., Syndulko K., Lindsley, D.B.: Electrocortical Signs of Levels of Processing: Perceptual Analysis and Recognition Memory. Psychophysiology, vol. 17, 1980, pp. 568-576.

48. Johnson, R., Pfefferbaum, A., Kopell, S.: P300 and Long-Term Memory: Latency Predicts Recognition Performance. Psychophysiology, vol. 22, 1985, pp. 497-507.

49. Donchin, E.: Surprise!...Surprise ? Psychophysiology, vol. 8, 1981, pp. 493-513.

50. Fabiani, M., Karis, D., Donchin, E.: P300 and Recall in an Incidental Memory Paradigm. Psychophysiology, vol. 23, 1986, pp. 298-308.

51. Karis, D., Fabiani, M., Donchin, E.: P300 and Memory: Individual Differences in the Von Restorff Effect. Cognitive Psychology, vol. 16, 1984, pp. 177-216.

52. Friedman, D., Sutton, S.: Event-Related Potentials During Continuous Recognition Memory. In Rohrbaugh, J.W., Johnson, R., Parasuraman, R. (Eds.), Eighth International Conference on Event-Related Potentials of the Brain: Research Reports. Stanford, 1986, pp. 227-229.

53. Goodin, D.S., Squires, K.C., Henderson, B.H., Starr, A.: Age-Related Variations in Evoked Potentials to Auditory Stimuli in Normal Subjects. Electroencephalography and Clinical Neurophysiology, vol. 44, 1978, pp. 447-458.

54. Syndulko, K., Hansch, E.C., Cohen, S.N., Pearce, J.W., Goldberg, Z., Montan, B., Tourtellotte, W.W. and Potvin, A.R.: Long-Latency Event-Related Potentials in Normal Aging and Dementia. In Courjon, J., Mauguire, F. and Revol, M. (Eds.), Clinical Applications of Evoked Potentials in Neurology. Raven Press: New York, 1982, pp. 278-285.

55. Pfefferbaum, A., Wenegrat, B.G., Ford, J.M., Roth, W.T., Kopell, B.S.: Clinical Application of the P3 Component of the Event-Related Potentials. II. Dementia, Depression and Schizophrenia. Electroencephalography and Clinical Neurophysiology, vol. 59, 1984, pp. 104-124.

56. Goodin, D.S., Squires, K.C., Starr, A.: Long Latency Event-Related Components of the Auditory Evoked Potential in Dementia. Brain, vol. 101, 1978, pp. 635-648.

57. Hansch, E.C., Syndulko, K., Cohen, S.N., Goldberg, Z.I., Potvin, A.R. and Tourtellotte, W.W.: Cognition in Parkinson Disease: An Event-Related Potential Perspective. Annals of Neurology, vol. 11, 1982, pp. 599-607.

58. O'Donnell, B.F., Squires, N.K., Martz, M.J., Chen, J.R.. Phay, A.J.: Evoked Potential Abnormalities and Neuropsychological Performance in Parkinson's disease. Biological Psychology, vol. 24, 1987, pp. 23-37.

59. Roth, W.T., Horvath, T.B., Pfefferbaum, A., Kopell, B.S.: Event-Related Potentials in Schizophrenics. Electroencephalography and Clinical Neurophysiology, vol. 48, 1980, pp. 127-139.

60. Verleger, R., Cohen, R.: Effects of Certainty, Modality Shift and Guess Outcome on Evoked Potentials and Reaction Times in Chronic Schizophrenics. Psychological Medicine, vol. 8, 1978, pp. 81-93.

61. Levit, R.A., Sutton, S. and Zubin, J.: Evoked Potential Correlates of Information Processing in Psychiatric Patients. Psychological Medicine, vol. 3, 1973, pp. 487-494.

62. Pass, H.L., Klorman, R., Salzman, L.F., Klein, R.H., Kasberg, G.B.: The Late Positive Component of the Evoked Response in Acute Schizophrenics During a Test of Sustained Attention. Biological Psychiatry, vol. 15, 1980, pp. 9-20.

63. Brecher, M., Porjesz, B., Begleiter, H.: The N2 Component of the Event-Related Potential in Schizophrenic Patients. Electroencephalography and Clinical Neurophysiology, vol. 66, 1987, pp. 369-375.

64. Duncan-Johnson, C.C., Roth, W., Koppell, B.S.: Effects of Stimulus Sequence on P300 and Reaction Time in Schizophrenics. In Karrer, R., Cohen, J., Tueting, P. (Eds.), Brain and Information: Event-Related Potentials. Annals of the New York Academy of Sciences, vol. 425, 1984, pp. 570-577.

65. Drake, M.E., Burgess, R.J., Gelety, T.J., Ford, C.E., Brown, M.E.: Long-Latency Auditory Event-Related Potentials in Epilepsy. Clinical Electroencephalography, vol. 17, 1986, pp. 10-13.

66. Callaway, E.: Human Information Processing: Some Effects of Methylphenidate, Age, and Scopolamine. Biological Psychology, vol. 19, 1984, pp. 649-662.

67. Halliday, R., Naylor, H., Callaway, E., Yano, L., Walton, D.: D-Amphetamine Speeds Both Stimulus and Response Processing. In Rohrbaugh, J.W., Johnson, R., Parasuraman, R. (Eds.), Eighth International Conference on Event-Related Potentials of the Brain: Research Reports. Stanford, 1986, pp. 93-95.

68. Hammond, E.J., Meador, K.J., Aung-Din, R., Wilder, B.J.: Cholinergic Modulation of Human Event-Related Potentials. Neurology, vol. 37, 1987, pp. 346-350.

69. Ruchkin, D.S., Sutton, S.: Equivocation and P300 Amplitude. In Otto, D. (Ed.), Multidisciplinary Perspectives in Event-Related Brain Potential Research. EPA-600/9-77-043. U.S.G.P.O.: Washington, D.C., 1978, pp. 175-177.

70. Johnson, R.: A Triardric Model of P300 Amplitude. Psychophysiology, vol. 23, 1986, pp. 367-384.

71. Donchin, E., Kramer, A., Wickens, C.: Applications of Brain Event-Related Potentials to Problems in Engineering Psychology. In Coles, M.G.H., Donchin, E., Porges, S.W. (Eds.), Psychophysiology: Systems, Processes, and Applications. New York: Guilford, 1986, pp. 702-718.

72. Rohrbaugh, J.W., Johnson, R., Parasuraman, R. (Eds.), Eighth International Conference on Event-Related Potentials of the Brain: Research Reports. Stanford, 1986.

73. Hillyard, S.A., Kutas, M.: Electrophysiology of Cognitive Processing. Annual Review of Psychology, vol. 34, 1983, pp. 33-61

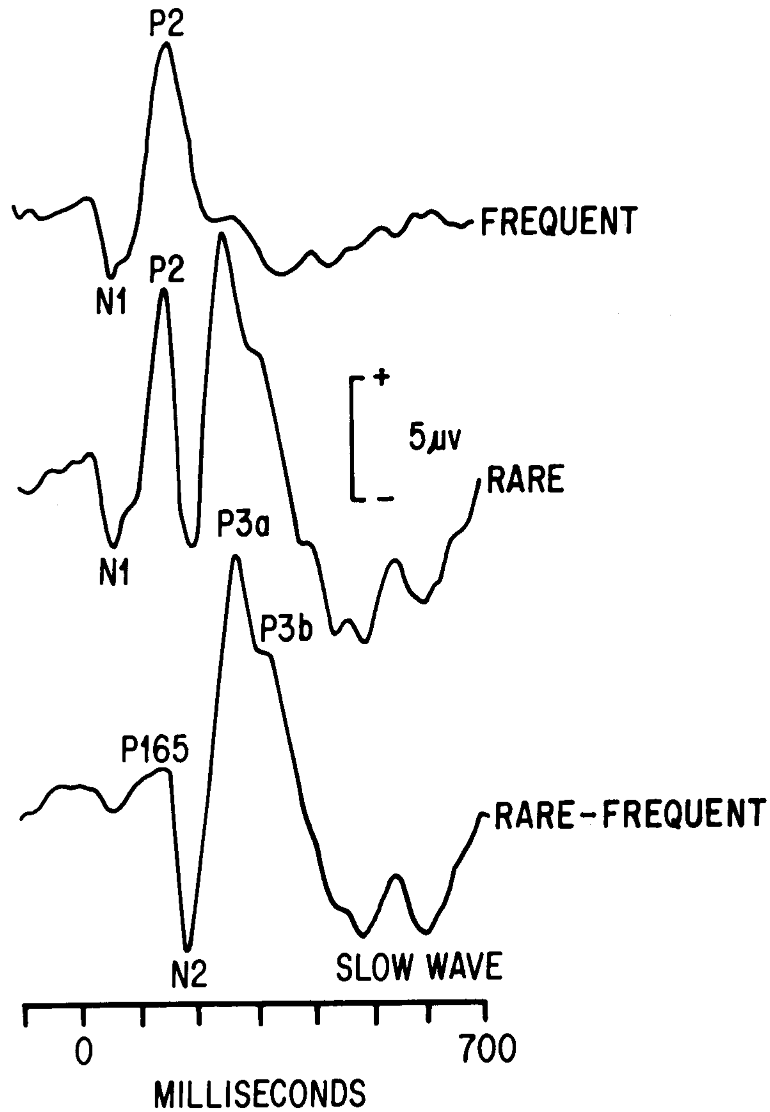


Figure 1. Evoked potentials averaged from frequent 1000 Hz tones and rare target 2000 Hz tones (probability = .10). Frequent tones elicit the N1-P2 components, while rare tones elicit both the N1-P2 and endogenous N2-P3 components. Subtraction of waveforms generated by rare tones from frequent tone waveforms isolates the endogenous components (P165, N2, P3a, P3b, and Slow Wave).

N88-23389 519-53

140807

25P.

S1195835

SPB

Processing Deficits in Monitoring Analog and Digital Displays: Implications for Attentional Theory and Mental-State Estimation Research

David G. Payne and Virginia A. L. Gunther

Department of Psychology and
Center for Cognitive and Psycholinguistic Sciences
State University of New York
University Center at Binghamton

Abstract

Subjects performed short term memory tasks, involving both spatial and verbal components, and a visual monitoring task involving either analog or digital display formats. These two tasks (memory vs. monitoring) were performed both singly and in conjunction. Contrary to expectations derived from multiple resource theories of attentional processes, there was no evidence that when the two tasks involved the same cognitive codes (i.e., either both spatial or both verbal/linguistic) there was more of a dual task performance decrement than when the two tasks employed different cognitive codes/processes. These results are discussed in terms of their implications for theories of attentional processes and also for research in mental state estimation.

Introduction

There has recently been considerable interest in assessing the patterns of interference effects obtained when operators simultaneously perform two or more tasks that require controlled information processing. It is commonly assumed that as the total amount of attention (or 'capacity' or 'mental resources') required to perform these tasks increases above some level, overall performance levels will decrease. This performance decrement is often assumed to follow the principle of graceful degradation outlined by Norman and Bobrow (ref. 1). Our research is directed towards the general goal of identifying performance deficits in dual-task situations involving tasks similar to those performed by operators in advanced flightdeck environments. Our interest, however, is not so much simply in the fact that performance in these situations falters when the operator is overloaded. Rather, we are primarily interested in determining the specific ways in which performance is affected when the total task demands exceed the limited information processing capabilities of the operator. For example, if a pilot cannot accurately read the information displayed on a CRT, what perceptual/cognitive processes are responsible for this performance decrement?

Due to the complexity of many of the tasks performed within the aerospace flight deck environment, there are many ways in which performance could be affected. If our goal is to determine how various mental states (e.g., boredom, fatigue) are related to performance within these complex environments,

then it is essential that we have an in-depth understanding of the factors that influence operators' behaviors in these situations. To foreshadow a bit, we would argue that the efforts to a) identify mental states using physiological indices and b) relate these mental states to performance in the flight deck environment can succeed only if we possess a concise knowledge of the cognitive processes affected by task demands.

The research in this article had several inter-related goals. The first was to attempt to determine the optimal format for presenting information to operators in a process control task. The process control task we employed exhibited two characteristics that make it similar to tasks performed by flightdeck personnel. First, there were a large number of display indicators that the subjects monitored. Second, although the subject was required to monitor all of the indicators, a response was required only when one of the indicator values exceeded the acceptable range. This latter task characteristic is analogous to when a pilot takes corrective action only when the actual airspeed deviates by a certain amount from the desired, or target, airspeed.

Our second goal was to examine how different types of display formats affect operators abilities to perform other ongoing activities. Towards this end we attempted to apply existing theories of attentional processes to predict performance levels in a dual task situation. Finally, we hoped that the results of this research would enable us to develop reasonable tasks for use in mental state estimation research.

We will review the information relevant to these three goals after first briefly describing the general approach taken in our research. To provide some insight into which factors affect performance in ongoing visual monitoring tasks, we employed a dual-task methodology (cf., ref. 2) that has proven useful to researchers investigating memorial and attentional processes in a variety of basic (e.g., refs. 3, 4, 5, 6) and applied (e.g., refs. 7, 8, 9) research settings. Although we describe the dual task method in detail when we present our main experiment, the basic logic behind this method is as follows. An operator is required to perform two tasks, both singly and in conjunction, with performance being measured in both the single and dual task conditions. One task is designated the primary task and the operator is instructed to attempt to maintain optimal performance on this task. Assuming that performing the two tasks concurrently exceeds the limited information processing capacity of the operator, performance levels on the secondary task can be used as an indirect estimate of the amount of capacity, or processing resources, required by the primary task. By varying the difficulty level of the primary and secondary tasks we can examine performance

across a range of performance conditions. (See refs. 10 and 11 for a detailed description of the application of the dual task methodology.) In addition, we can investigate how different versions of these tasks fare when performed in conjunction with other tasks from real-world multi-task situations.

One final point regarding our general research strategy. We made a deliberate attempt in our study to investigate theoretically important issues using tasks that have relevance to performance in real world situations. We believe that as a general research strategy this approach helps to increase the applicability of the research (and thus aids the human factors specialist), and also allows the basic researcher to address theoretical issues under highly controlled laboratory conditions.

Attentional Limitations in Performing Controlled Information Processing Tasks

Our research relied heavily upon current models and theories of human attentional processes. In this section we briefly review these models and theories. The reader should note that this review is not intended to be inclusive, as several excellent reviews exist in the literature (e.g., refs. 10 and 12). (Readers already familiar with modern theories of attention can go directly to the descriptions of the present research.)

It almost goes without saying that in everyday life people are often engaged in tasks that require them to perform two or more functions simultaneously (e.g., driving a car while attempting to locate a specified street address). The literature on attentional processes and information processing is replete with cases in which human performance suffers when a person is required to perform two or more tasks concurrently (e.g., refs. 13, 14). There are also cases in which such time-sharing is carried out quite efficiently (e.g., refs. 15, 16). One of the puzzles facing theorists and researchers over the last 20 to 30 years has been to specify under what conditions two tasks may be time-shared efficiently (e.g., walking while talking) and under what other conditions time sharing is inefficient (e.g., carrying on a conversation while reading).

Historically, there have been two general approaches towards providing a theoretical explication of such time-sharing phenomena. In the 1950s and 1960s there were a number of investigations showing that humans were extremely limited in their ability to attend to two separate auditory messages (e.g., refs. 17, 13, 14). Findings such as these lead to the development of structural theories (e.g., refs. 17, 3, 18) that attempted to identify at which point in the processing

of information did the "bottleneck" occur that seemed to limit performance in dichotic listening experiments, as well as in other cases in which people showed limitations in their ability to process information efficiently (e.g., the psychological refractory period phenomena; see ref. 19 for a review). According to these structrural theories, then, the degraded performance one observes when the operator attempts to process large amounts of information is attributable to the manner in which the information processing stages are "structured" or configured.

An alternative approach to explaining time-sharing was offered by the capacity theories proposed in the 1960s and 1970s (e.g., refs. 20, 21). This approach is best exemplified by Kahneman's theory in which he proposed that there exists a single, limited "pool" of capacity that can be allocated to performing all ongoing controlled information processing tasks. According to this view, the limitation in time sharing is not one of limited access to processing structures, but rather it is that the processing structures can only function when "capacity" is allocated to those structures. The efficiency with which two tasks may be time shared depends upon the availibility of sufficient capacity to perform the necessary information processing. If there is adequate capacity to meet the demands of the two tasks, then these tasks may be performed as efficiently in conjunction as they can be performed singly; if the total capacity required by the two tasks exceeds the "pool" of available capacity, then performance in the dual task condition will fall below what is observed in the single task conditions.

Although both structural and capacity theories are capable of explaining a great deal of the data on time sharing, there are numerous findings that indicate that these theoretical conceptualizations are too impoverished to provide a complete explication of the phenomena of interest. (For a review of these difficulties, see refs. 22 and 12.) As a result, there has recently been proposed a third approach to time sharing, namely resource theory (e.g., ref. 22). Resource theory has been successfully applied in a number of investigations, including basic research (e.g., refs. 23, 24, 25) and applied human factors research (e.g., refs. 7, 26). This approach to understanding human cognitive abilities appears to have great promise, although there have been some arguments made against theories that propose the existence of multiple resources (e.g., ref. 27). Since our research utilizes a resource theory approach, we will describe the general concepts embodied in multiple resource theory in some detail.

Navon and Gopher (ref. 22) proposed that instead of a single pool of capacity that may be shared among various processing structures, it might be better to envision the human

cognitive system as being comprised of a limited number of processing "resources". Capacity and resources are both hypothetical constructs that are used to refer to underlying commodities that enable a person to perform some task(s). A major difference between the concepts of capacity and resources is that capacity is generally assumed to be rather amorphous, in the sense that it may be allocated to any processing stage or structure, whereas resources are less general in nature. That is, it is assumed that resources may only be allocated to specified processes or subprocesses. It is further assumed that several types of resources exist and these differ in kind, such that they may not be readily substituted for one another. (Multiple resource theories do allow for some substitution of resources. However, there is generally a loss of processing efficiency associated with these substitutions; we will return to the issue of processing efficiency momentarily.)

Recall that capacity theory assumed that a) there was a single pool of capacity, and b) that, in a dual task situation, if there were spare capacity left from performing Task A, then that "spare" capacity could be allocated to performing Task B. Multiple resource theory, on the other hand, suggests that if Task B requires a particular resource that is in short supply, then even if other resources are readily available (e.g., those resources not required to perform Task A), these other resources can not be utilized efficiently in performing Task B.

As mentioned previously, multiple resource theory assumes that differing resources are differentially efficient when applied to processes or subprocesses. Efficiency here is used in the econometric sense of marginal efficiency (i.e., the change in performance level observed when one unit of a resource is added to or removed from a process). Finally, different tasks require differing resources for the processing involved in that task to be completed. The resources required to perform a task is generally referred to as that task's resource composition.

To summarize according to multiple resource theories, the following factors are assumed to affect performance in single and dual task situations: (a) the resource composition(s) of the task(s) under investigation, (b) the amount of each resource type available to be allocated to the task(s), and (c) the relative efficiency of the resources allocated to the task(s). One obvious difficulty with an unconstrained multiple resource model is the issue of how one determines <u>a priori</u> precisely what constitutes a resource and which of these putative resources are required to perform specified tasks. Without appropriate limitations, resource theory could follow in the path of instinct theory and faculty psychology and propose resources <u>ad infinitum</u>. There are

however, two promising approaches for limiting the number and type of resources incorporated in the models.

One approach to the problem of identifying resources is to view each cerebral hemisphere as having its own processing resources. This perspective draws heavily upon findings indicating that the two hemispheres are specialized for performing different functions (e.g., spatial tasks are assumed to rely upon right hemisphere resources, verbal tasks are assumed to rely upon left hemisphere resources). There is considerable empirical support for this general approach to resource theory (e.g., refs. 23, 24, 28, 29, 30, 31, 26,32).

A second approach to attempting to limit the proliferation of processing resources is best exemplified by the work of Wickens (ref. 33). This approach examines the types of tasks that produce interference effects when performed in conjunction and then uses these data to discern the specific types of tasks that utilize similar resources. The general underlying assumption here is that if two tasks interfere with one another when performed in conjunction, then these tasks must employ the same or similar resources; if there is little or no dual-task interference then the resource compositions of the two tasks overlap only minimally.

Using this approach, Wickens (refs. 33, 12) has identified the following as candidates for processing resources: (a) the type of input and output modality (e.g., visual vs. auditory stimuli; manual vs. vocal responses), (b) the code or representational format utilized by the subject (e.g., a verbal/linguistic code vs. a spatial code), (c) the stage of processing (e.g., encoding, central processing and response selection, response execution), and (d) the hemisphere of processing (cf. the distinctions noted above in the first approach). The present research employed the distinction between verbal/linguistic codes vs. spatial codes in an effort to apply multiple resource theory to a real world information processing task.

Application of Attentional Theory to a Visual Monitoring Task

As indicated previously, our research is couched within the framework provided by multiple resource theory. One of our major goals was to examine the patterns of interference effects obtained in dual task conditions when subjects perform visual monitoring tasks. According to multiple resource theory, the pattern of performance observed in a dual task situation depends upon the resource composition of the primary and secondary tasks. According to this view, then, it is possible for two tasks that have very different resource compositions to show different levels of dual task performance as a function of the secondary task with which they are conjoined. That is, a

task that has a large spatial processing component may produce large dual task performance decrements when conjoined with a secondary task that also utilizes spatial codes but shows little or no dual task decrement when conjoined with a secondary task that utilizes verbal/linguistic codes.

If the concept of multiple resources (as defined by the nature of the codes involved in the processing tasks) is accurate, then this has implications for the design of displays for person-machine systems. For example, if an operator is performing a series of tasks that are highly spatial in nature (e.g., flying an aircraft), then the use of displays that rely heavily upon spatial processes may not be optimal. In this case it may be better to use displays that require verbal/linguistic processes. To test this hypothesis, we employed a laboratory analog of a process control task originally described by Hanson, Payne, Shively and Kantowitz (ref. 9).

Hanson et al (ref. 9, Experiment 2) required subjects to monitor either an analog or a digital display presented on a cathode ray tube (CRT). In both display formats there were indicators that presented data corresponding to the constantly varying outputs of a simulated process control system. The subject's task was to monitor the system outputs and take a corrrective action whenever one of the displays went beyond a specified range. In the analog condition the system output values were represented by the length of the lines in a display similar to a histogram. In the digital display condition the actual numerical value of each system variable was presented. Coupled with this visual monitoring task was either a 2- or 4-choice auditory choice reaction time task. These reaction time tasks were included in order to assess the processing demands of the analog vs. digital displays. Results showed that increasing the difficulty level (operationalized as the number of display indicators presented) of the analog displays had little effect on performance in the auditory choice reaction time task but had a sizable impact on performance when subjects were monitoring the digital displays. Hanson et al interpreted their results within a single capacity framework, arguing the the analog task required less capacity to perform and this then resulted in less performance decrement as the secondary task difficulty was increased.

Our research was designed as a follow-up to the study by Hanson et al (ref. 9). We presented subjects with two tasks, a short term memory task and either a digital or an analog visual monitoring task similar to those used by Hanson et al. For both of these tasks (memory and monitoring), we constructed one version of the task that relied predominately upon spatial codes/processes and a second that relied upon verbal/linguistic codes/processes.

Our first experiment was a pilot study designed to establish appropriate task parameters for the memory task in the main experiment. This pilot study also provided information regarding the processing requirements of the short term memory tasks. In the pilot study subjects viewed a computer monitor containing a four x four (16 cell) matrix. Three letter English words were presented one at a time within single cells of the matrix using a three sec presentation rate and a 1 sec interstimulus interval. For different trials, the instructions for the memory task were intended to tap either spatial processing, verbal/linguistic processing, or a combination of these two types of processes.

Across trials, subjects were presented with lists of varying length (range = 4 - 9 items) and were given one of three recall tasks. In the item condition, subjects were instructed to recall only the items from the target list. On the location trials the subjects' task was to remember the locations within the matrix that contained items during the list presentation. Finally, in the item + location + order condition subjects were required to place the items they recalled in the correct locations within the matrix and also indicate the serial order with which these items appeared in the list. We assumed that the item task loaded primarily upon verbal/linguistic codes (or processes), the location task loaded primarily upon spatial codes/processes, and that the item + location + order task tapped both types of processes.

In addition to studying the lists, subjects in the pilot experiment were also given one of three tasks to perform between the end of list presentation and the start of the recall test. In the spatial interpolated task subjects were presented with pairs of symbols (e.g., ####, &&&&) in different locations on the CRT screen and were asked to decide if these items were in certain spatial arrangements (e.g., 'Is the #### above the &&&&?'). These items appeared in sequential pairs, with the direction corresponding to the above or below decision being indicated before the two comparison stimuli were presented. Subjects indicated their decisions by pressing buttons on a response box in front of the CRT. The second interpolated task was a numerical decision task analogous to the spatial task. Subjects were given a 'direction' (greater than, less than), followed by two successive three-digit numbers. The subjects' task was to decide if the two items were in the designated numerical relations. Finally, in the Brown-Peterson task subjects were given a three digit number and asked to count backwards out load by threes from that number as rapidly as possible. Each of these tasks lasted for 60 sec with the recall tests being given immediately after the interpolated tasks.

The results of this pilot study indicated that, not surprisingly, recall performance was affected by list length, with more items being recalled as list length was increased. More importantly, the comparison of the several combinations of recall task x interpolated activity offered support for the notion that the item and location recall tasks were differentially affected by the interpolated tasks. First, the Brown-Peterson task, which requires subjects to keep a mental tally of the current numeric item, subtract 3 from that item and then repeat the entire process over again, produced the lowest recall levels of any of the three interpolated tasks. Also, there was some evidence that the item and location recall tasks were diffentially affected by the spatial and numerical interpolated tasks. Finally, the item + location + order condition produced far lower performance levels than the other two recall condition.

Taken together then, these pilot results indicate that the memory task is sensitive to the memory load; the item + location + order condition imposed the greatest memory load and also produced the lowest recall levels. Furthermore, the spatial, numerical, and Brown-Peterson tasks produced differential degrees of within-trial interference in the item, location, and item + location + order conditions. This latter finding supports our conjectures about the codes/processes involved in these memory tasks. Finally, the results of the pilot study indicated that, for the stimulus items and presentation conditions used in the main experiment, a six item list would produce performance levels in the range of 50% to 95% correct recall, depending upon the recall task. With these findings in hand we proceeded to the main experiment.

Method

Subjects and Design. Eighteen male and 18 female undergraduates at SUNY - Binghamton participated in partial fulfillment of a course requirement for research experience or library research. Of each same-sex group of 18 subjects, 6 were left handed and 12 were right handed, with handedness being determined by subjects self-report and preferred writing hand.

Subjects participated in three 9-trial blocks, two single task blocks (Blocks 1 and 3) and one dual task block (Block 2). In the single trial blocks there were six memory task trials followed by three visual monitoring task trials. In these single trial blocks order of presentation of the three types of memory tasks (item, location, and item + location + order) was counterbalanced across subjects such that each subject received one of each type of memory task in trials 1 - 3 and a different order of these three memory tasks in trials 4 - 6. Each trial consisted of a different set of 6 items and across subjects the same items were presented on each trial and thus each set of

six memory items/locations appeared equally often in each memory condition. Following the six memory task trials there were three visual monitoring trials. Blocks 1 and 3 were identical, with the exception that a different set of memory items was used in each block.

In Block 2 subjects were presented with nine trials in which they performed both the memory task and the visual monitoring task. The nine trials were broken into three sets of three trials each. Each of the three triads contained one of each of the three memory tasks (i.e., item, location, and item + location + order). Across the three triads the order of memory tasks within a triad was counterbalanced using a Latin square design.

One half of the subjects (nine males and nine females) performed a digital visual monitoring task and the remaining subjects performed an analog monitoring task. Within each set of nine same-sex subjects assigned to each type of monitoring task, three were left handed and six were right handed. Thus the between subjects factors in this experiment were type-of-visual-monitoring task (analog vs. digital), sex, and handedness. (These latter two subject variables were included to address issues unrelated to the primary goals of the present study and hence will not be described any further in this report.) The within subjects factors were type of trial (single task vs. dual task) and type of memory task (item, location, item + location + order) on the single task memory trials (trials 1 - 6 of Blocks 1 and 3) and dual task trials (Block 2).

Procedure. Both the short term memory task and the visual monitoring task were controlled by an Apple IIe microcomputer equipped with an Apple color monitor, a millisecond timer and an eight key response box. For the short term memory task subjects viewed a 16 cell (4 x 4) matrix on the computer monitor. A trial consisted of presenting 6 three letter English words, with each word appearing in a different, randomly determined location within the 16 cell matrix. Words were presented at a three sec presentation rate with a one sec interstimulus interval. The same presentation format was used with each of the three memory tasks, with the sole difference between tasks being the instructions given to subjects prior to the trial and the corresponding differences in the retention measures. For the item trials subjects were given standard free recall instructions indicating that their task was to study the items so that they could recall the items from the study list in any order they choose. On the location trials subjects were told that they were not responsible for remembering the actual items that were presented but rather they would be asked to recall which of the cells contained a word during the list presentation. For item + location + order trials subjects were

told that they were to try to remember the items, the locations within the matrix that each item appeared and also the serial presentation order (i.e., first, second, ... sixth) of the items. After these instructions were given subjects were presented with the six target items for that trial. In the single task item trials, after the list was presented, subjects wrote the target items on a sheet of blank paper. In the item and item + location + order conditions, after the list was presented subjects were given a sheet of paper with a 4 x 4 matrix printed on it and were asked to recall the information that they had been instructed to memorize on that trial. On location trials subjects were asked to place an X in each cell of the matrix in which a word had appeared during the list presentation. For item + location + order trials subjects were told to write the items in the cells in which they had appeared and also indicate the order of appearance by numbering the cells from 1 to 6. Subjects were given as much time as needed to complete the tests.

In the single task visual monitoring trials subjects viewed either an analog or a digital display. Both types of displays presented eight indicators representing the status of simulated system outputs. The subject's task was to monitor the eight indicators and "reset" any indicator (by pressing a button on the response box) that exceeded preset boundaries. For the digital displays, the value of each indicator was presented in the center of a box and the upper (282) and lower (110) limits for these indicators were printed above and below the box containing the indicator value (see Figure 1). At the onset of the trial, each indicator started near the middle of the range of acceptable values and began either consistently increasing or decreasing. The software that controlled the monitoring task "updated" each indicator in succession, recorded when the indicator value first exceeded the upper or lower boundary, when the subject "reset" each indicator, and also any "resets" that the subject attempted before the indicator had exceeded its boundary. Once the trial began, each indicator continued to either increase or decrease, with the magnitude of each change being a value chosen at random from the range ± 1 to ± 20 units. After an indicator reached its maximal (187) or minimal (105) value, the indicator no longer changed until it was reset by the subject pressing the button corresponding to that indicator. (Each button was associated with a single indicator in a consistent 1 to 1 mapping.) Once an indicator was reset it was then restarted at a value close to the middle of the range and began changing again, either increasing or decreasing. The direction of change was random across resets, thus after a reset an indicator could change in the same direction as it had been previously or it could move in the opposite direction.

The analog monitoring task was identical to the digital task in all regards save the manner in which the indicators were presented. (See Figure 2.) The same algorithm was used to determine the rate and direction of change of each indicator, only now the values were used to plot analog representations of these values, with increasing values moving upwards and decreasing values moving downwards. The rates of updating and changing the displays were held constant across the two display types.

For both types of single task visual monitoring trials subjects performed the monitoring task for 60 seconds. The parameters of this task were such that, on average, approximately 35 - 40 indicators would require resetting during the trial if the indicators were reset immediately upon crossing the boundaries. For the dual task trials (Block 2) subjects were first given the target items to study, followed by one min of visual monitoring and then the recall test for the memory task information. The end of the memory task list presentation was followed immediately by the start of the monitoring task, with the only delay being the time needed for the computer to generate the monitoring displays.

Results and Discussion

Performance in the single and dual task trials was evaluated using several dependent variables. For the item recall and location recall condition subjects were given credit for correctly recalling the target information. In the item + location + order condition, performance was measured by scoring both the number of items correctly recalled and the number of locations correctly recalled. For the location measure subjects were given credit for recalling the item's location only if the correct item also appeared in that location. For the analog and digital visual monitoring task we measured the mean reaction time for resetting the indicators and the mean number of errors made per trial, with an error being operationally defined as attempting to reset an indicator before it reached its boundary.

Presented in Table 1 are the mean recall rates for the two single task trial blocks. As expected, on the single task recall trials there were no differences in the performance levels between the analog and digital groups for any of the recall measures (all $\underline{p}$s $> .20$). Consistent with the pilot study, the recall rates for the item and location information was significantly better in the item and location conditions than in the item + location + order conditions, and this held for both the analog and digital groups ($\underline{p} < .05$). (All effects called significant were assessed using appropriate statistical measures and had $\underline{p}$ values $< .05$.) This finding indicates that there were different levels of difficulty across the three

recall tasks, with the two tasks requiring memory for a single type of information (i.e., the item and location conditions) producing better performance than the condition that required subjects to retain several different types of information (i.e., the item + location + order condition). Thus subjects in the two visual monitoring groups were performing at an equivalent level on the single task recall trials and the item and location recall tasks produced better performance levels than the item + location + order task.

An important point to note with regard to the recall data is that the performance levels were stable across the two blocks of trials. None of the recall conditions showing a significant change in mean correct recall from Block 1 to Block 3. Furthermore, this stability in performance levels is not simply due to a ceiling effect in the item and location conditions: Performance levels in the item + location + order condition were at approximately 70% correct recall. Despite there being considerable room for an improvement in recall, there was no evidence of a change in performance levels across the session.

Finally, although the performance levels on the item trials was numerically greater than that obtained on the location trials, this difference was not significant (1p0 > .10). This suggests that when subjects were only required to perform the memory task, they produced equivalent performance levels in the tasks designed to tap either spatial processing (i.e., the location condition) or verbal/linguistic processing (i.e., the item condition). This suggests, then, that these two tasks are roughly equivalent in terms of their "difficulty".

The results from the single task visual monitoring trials are presented in Table 2. Replicating Hanson et al (ref. 9), the digital task produced significantly longer reaction times than the analog task. There were also significant differences in the errors rates across these two conditions, with the analog condition producing the higher error rate. The differences in reaction times and error rates would seem to indicate that these results represent a classic case of a simple speed-accuracy tradeoff. However, observations of subjects performing these tasks, as well as subjects' introspective self reports, suggest that this was not the case in the present experiment.

Recall that in this task an error corresponds to the subject attempting to reset an indicator prior to its crossing the boundary. Subjects in the analog condition seemed to be making errors because they were attempting to "predict" when an indicator would cross the boundary. However, because the magnitude of the increment/decrement on each update of an indicator was random, these predictions could not be 100%

accurate. Thus as a result of using this prediction strategy subjects occasionally attempted to reset an indicator before it had crossed the boundary. Note, however, that the use of this prediction strategy requires that subjects selectively attend to the indicators that were nearing the threshold for resetting. This selective attention strategy is possible only if the subjects were efficient at monitoring the relative positions of all eight indicators.

In contrast to the analog condition, subjects in the digital condition were quite slow in resetting the indicators. Furthermore, these subjects did not make many "prediction" errors. This low error rate seems to be due to the fact that subjects were unable to efficiently discern which indicators were nearing the boundaries. Subjects in the digital condition did not appear to be able to focus attention on the indicators that were nearing the boundaries and hence they produced long reaction times and low error rates.

Another aspect of the single task reaction times that warrants notice is the fact that subjects' reaction times continued to improve across the session. This suggests that subjects had not reached asymptotic performance levels and thus the processes involved in monitoring the displays had not become "automatic" processes. Based on the distinction of automatic vs. controlled processes (cf. refs. 34, 35) the visual monitoring task still required capacity/resources for its completion. To determine the nature and extent of the capacity/resources required to perform these tasks we need to examine the performance levels in the dual task trials from Block 2.

The mean recall levels for the dual task trials are presented in Table 3. These data indicate that the recall levels in the dual task trials were very similar to those observed in the single task trials (see Table 1). This suggests that subjects were allocating sufficient capacity/resources to the memory task in the dual task trials so as to maintain dual task performance at the level of the single task trials.

A second interesting aspect of the dual task recall data is that there was no evidence of selective interference between the analog and digital monitoring tasks and the three types of recall task. That is, while there were significant differences between the item and location conditions vs. the item + location + order condition, the differences were of approximately the same magnitude for the two types of monitoring tasks. This lack of a memory task x visual monitoring task interaction raises the issue of whether, as predicted by some multiple resource models, there was selective interference in the performance levels of the visual monitoring tasks.

The mean reaction times and error rates for the visual monitoring dual task trials are presented in Table 4. As in the single task trials, there was a significant main effect of visual monitoring condition in both the reaction time data and the error rate data. The analog condition produced shorter reaction times and higher error rates. More importantly, however, there was no evidence that performance on either of these tasks was affected by the type of information subjects had encoded prior to beginning the visual monitoring task. Although the results of the pilot study indicated that performing the item and location memory tasks requires the use of verbal and spatial codes, respectively, there was no indication that maintaining these codes in short term memory interfered with performance on the analog and digital visual monitoring task. This finding offers no support for the notion of separate processing resources corresponding to verbal/linguistic and spatial codes or processes.

General Discussion

One of the goals of this study was to examine the relative difficulty of monitoring analog and digital displays. The results of the present experiment are consistent with those reported by Hanson et al (ref. 9) demonstrating that analog displays are monitored more efficiently than are digital displays. One question that can be asked of these findings is the extent to which they generalize to trained pilots performing actual flight operations. The results of a recent study by Koonce, Gold, and Moroze (ref. 36) indicate that the analog superiority obtained with college students performing our laboratory task is also obtained when both college students and pilots "fly" a flight deck simulator. Koonce et al had flight naive and experienced pilots perform basic flight maneuvers using either analog or digital displays. They found that for both subject populations the analog displays resulted in superior performance to the digital displays. Thus three separate studies provide converging evidence that analog displays are monitored more efficiently than are digital displays.

A second goal of our study was to examine the attentional requirements of monitoring the analog and digital displays. Recall that Hanson et al (ref. 9) used visual monitoring tasks similar to those used in the present study. Those researchers examined the amount of capacity required to monitor the two types of displays by using a nonverbal, auditory secondary task. Koonce et al included a condition in which subjects "flew" the simulator while also performing an aural secondary task (detecting specified patterns of digits). Using these online secondary tasks, both studies found evidence of better secondary task performance with the analog displays than the digital displays. This suggests that when auditory, online

secondary tasks are used there is a difference in secondary task performance as a function of the type of visual display employed.

In the present experiment we employed a memory preload technique to assess the capacity/resource demands of the visual monitoring task. This secondary task required subjects to maintain different types of cognitive codes in short term memory for the duration of the visual monitoring task. Under these conditions we found no evidence of a difference in secondary (or primary) task performance as a function of the specific type of primary and secondary tasks. One of the questions that remains to be answered is why different patterns of secondary task performance were obtained in these three studies.

There are several differences between the procedures used by Hanson et al and those employed in the present experiment, and even greater procedural variations between the study of Koonce et al and our experiment. Based on the available data it is not possible to identify the precise cause of the different patterns of secondary task results. One possible explanation is that perhaps the modality of the secondary task is crucial (we used a visual task whereas Hanson et al and Koonce et al used an auditory task). Alternatively, perhaps the online and preload techniques are not equivalent in the extent and nature of the information processing load they impose upon the subjects. Research ongoing in our laboratory is attempting to resolve these and other issues related to the general goal of providing an accurate characterization of the attentional demands of various visual and auditory information processing tasks.

Finally, in keeping with the goals of the Mental State Estimation Workshop, there are two additional points that we would like to make. The first concerns the implications of the present study for attentional theory. It is important to note that our study was designed to test one instantiation of a multiple resource model, namely a model that postulates different resources for spatial and verbal/linguistic processes or codes. Although our results provide no evidence for this model, it would be premature to discard either the specific multiple resource model we tested or the more general theoretical concept of multiple resources. In terms of the specific model, it is possible that our procedures simply did not stress the subjects information processing system sufficiently to produce the selective interference predicted by the spatial vs. linguistic distinction. Regarding the general theory, it is possible that there are in fact multiple resources, but that the spatial vs. linguistic dimension is not one of the bases for these different processing resources.

The second point we would like to make concerns mental state estimation research. We believe that in order for researchers to relate mental states (as indexed by physiological indices obtained while subjects are engaged in cognitively demanding tasks) to behavior (i.e., the performance observed on these tasks), it is essential that the investigators fully understand the cognitive processes operating when subjects perform these tasks. Mental state estimation researchers and investigators interested in developing models and theories of human information processing could both profit from collaborative research aimed at relating mental states, cognitive processes, and behavior. Such a collaborative, interdisciplinary approach will greatly help to advance our understanding of how people perform various real world tasks of interest.

This research was supported in part by a BRSG Grant S07RR07149-12 awarded by the Biomedical Research Support Grant Program Division of Research Resources, National Institutes of Health.

REFERENCES

1. Norman, D., & Bobrow, D. (1987). On data-limited and resource-limited processes. Cognitive Psychology, 7, 44-64.

2. Ogden, G. D., Levine, J. M., & Eisner, E. J. (1979). Measurement of workload by secondary tasks. Human Factors, 21, 529-548.

3. Kerr, B. (1973). Processing demands during mental operations. Memory and Cognition, 1, 301-412.

4. Proctor, R. W., & Proctor, J. D. (1979). Secondary task modality, expectancy, and the measurement of attentional capacity. Journal of Experimental Psychology: Human Perception and Performance, 5, 610-624.

5. Roediger, H. L., Knight, J. L., & Kantowitz, B. H. (1977). Infering decay in short-term memory: The issue of capacity. Memory and Cognition, 5, 167-176.

6. Tyler, S., Hertel, P., McCallum, M., & Ellis, H. (1979). Cognitive effort and memory. Journal of Experimental Psychology: Human Learning and Memory, 5, 607-617.

7. Carswell, C. M., & Wickens, C. D. (1985). Lateral task segregation and the task-hemispheric integrity effect. Human Factors, 27, 695-700.

8. Damon, D. (1985). The relation between Type A behavior pattern, pacing, and subjective workload under single- and dual-task conditions. Human Factors, 27, 675-680.

9. Hanson, R. H., Payne, D. G., Shively, R. J., & Kantowitz, B. H. (1981). Process control simulation research in monitoring analog and digital displays. Proceedings of the Human Factors Society 25th Annual Meeting, 154-158.

10. Kantowitz, B. H. (1985). Channels and stages in human information processing: A limited analysis of theory and methodology. Journal of Mathematical Psychology, 29, 135-174.

11. Schweickert, R., & Boggs, G. J. (1984). Models of central capacity and concurrence. Journal of Mathematical Psychology, 28, 223-281.

12. Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), Varieties of Attention. New York: Academic Press.

13. Cherry, C. (1953). Some experiments on the recognition of speech with one and two ears. Journal of the Acoustical Society of America, 23, 915-919.

14. Treisman, A. M. (1964). Selective attention in man. British Medical Bulletin, 20, 12-16.

15. Allport, D. A., Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. Quarterly Journal of Experimental Psychology, 24, 225-235.

16. Spelke, C., Hirst, W., & Neisser, U. (1976). Skills of divided attention. Cognition, 4, 215-230.

17. Broadbent, D. (1958). Perception and communication. Oxford: Pergamon.

18. Norman, D. (1968). Toward a theory of memory and attention. Psychological Review, 75, 522-536.

19. Kantowitz, B. H. (1974). Double stimulation. In B. H. Kantowitz (Ed.), Human information processing: Tutorials in performance and cognition. Hillsdale, N.J.: Erlbaum.

20. Kahneman, D. (1973). Attention and effort. Englewood Cliffs, N.J.: Prentice-Hall.

21. Moray, N. (1967). Where is capacity limited: A survey and a model. Acta Psychologica, 27, 84-92.

22. Navon, D., & Gopher, D. (1979). On the economy of the human information. processing system. Psychological Review, 86, 214-255.

23. Friedman, A., & Polson, M. C. (1981). The hemispheres as independent resource systems: Limited-capacity processing and cerebral specialization. Journal of Experimental Psychology: Human Perception and Performance, 7, 1031-1058.

24. Friedman, A. P., Campbell, M., Dafoe, C. G., & Gaskill, S. S. (1982). Divided attention within and between hemispheres: Testing a multiple resources approach to limited-capacity information processing. Journal of Experimental Psychology: Human Perception and Performance, 8, 625-650.

25. Gopher, D., Brickner, M., & Navon, D. (1982). Different difficulty manipulations interact differently with task emphasis: Evidence for multiple resources. Journal of Experimental Psychology: Human Perception and Performance, 8, 146-157.

26. Wickens, C. D., Vidulich, M., & Sandry-Garza, D. (1984). Principles of S-C-R compatability with spatial and verbal tasks: The role of display-control location and voice interactive display-control interfacing. Human Factors, 26, 533-543.

27. Navon, D. (1984). Resources - a theoretical soup stone? Psychological Review, 91, 216-234.

28. Hellige, J. B., Cox, P. J., & Litvac, L. (1979). Information processing in the cerebral hemispheres: Selective hemispheric activation and capacity limitations. Journal of Experimental Psychology: General, 108, 251-279.

29. Kinsbourne, M., & Hicks, R. (1978). Functional cerebral space. In J. Requin (Ed.), Attention and Performance VII. Hillsdale, N. J.: Erlbaum.

30. Moscovitch, M., & Klein, D. (1980). Material-specific perception for visual words and faces. Journal of Experimental Psychology: Human Perception and Performance, 6, 590-603.

31. Pritchard, W. S., & Hendrickson, R. (1985). The structure of human attention: Evidence for separate spatial and verbal resource pools. Bulletin of the Psychonomic Society, 23, 177-180.

32. Wickens, C. D., & Sandry, D. L. (1982). Task-hemisphere integrity and dual-task performance. Acta Psychologia, 52, 2-20.

33. Wickens, C. D. (1980). The structure of attentional resources. In R. S. Nickerson (Ed.), Attention and Performance VII. Hillsdale, N. J.: Erlbaum.

34. Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. Psychological Review, 84, 1-66.

35. Shiffrin, R. M., & Schneider, W. C. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.

36. Koonce, J. M., Gold, M., & Moroze, M. (1986). Comparison of novice and experienced pilots using analog and digital flight displays. Aviation, Space, and Environmental Medicine, 57, 1181-1184.

Table 1

Mean Recall Levels for the Item, Location and Item + Location + Order Trials in the Single Task Conditions of Blocks 1 and 3

Group	Item	Location	Item + Location + Order: Item Scoring	Item + Location + Order: Location Scoring
		Block 1		
Analog Monitoring	5.36	5.50	4.47	3.97
Digital Monitoring	5.25	5.64	4.58	4.22
Mean	5.31	5.57	4.53	4.10
		Block 3		
Analog Monitoring	5.33	5.72	4.56	3.92
Digital Monitoring	5.36	5.64	4.75	4.39
Mean	5.35	5.68	4.65	4.15

(Column heading spanning all recall columns: Recall Condition)

Table 2

Mean Reaction Time (RT) and Error Rates for the Analog and Digital Monitoring Groups in the Single Task Trials of Blocks 1 and 3

Group	RT (in sec.)	Error Rate
	Block 1	
Analog Monitoring	2.59	9.11
Digital Monitoring	5.29	2.80
	Block 3	
Analog Monitoring	2.10	5.93
Digital Monitoring	3.41	3.72

Table 3

Mean Recall Levels for the Item, Location, and Item + Location + Order Trials in the Dual Task Conditions of Block 2

Group	Item	Location	Item + Location + Order: Item Scoring	Item + Location + Order: Location Scoring
Analog Monitoring	5.37	5.22	4.76	4.20
Digital Monitoring	5.41	5.59	4.67	4.29

(Recall Condition spans Item, Location, and Item + Location + Order columns.)

Table 4

Mean Reaction Time (RT) and Error Rate for the Analog and Digital Monitoring Tasks in the Dual Task Trials of Block 2

	Recall Task		
Group	Item	Location	Item + Location + Order
Analog Monitoring			
RT (in Sec.)	2.34	2.39	2.31
Error Rate	5.68	5.55	6.59
Digital Monitoring			
RT (in Sec.)	3.85	3.72	3.94
Error Rate	3.52	3.17	3.72

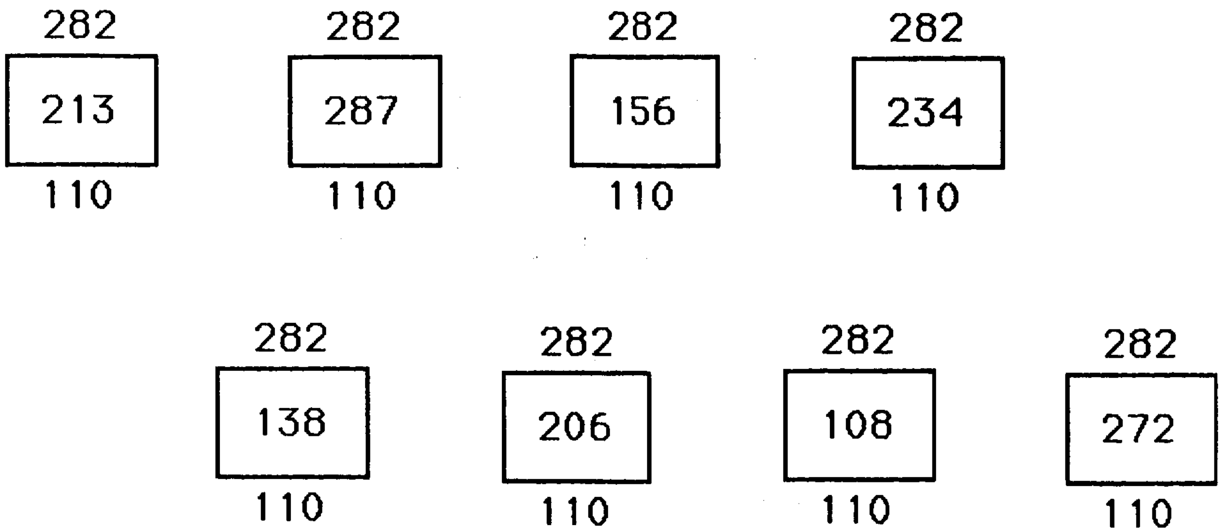


Figure 1. Example CRT Display for the Digital Visual Monitoring Task.

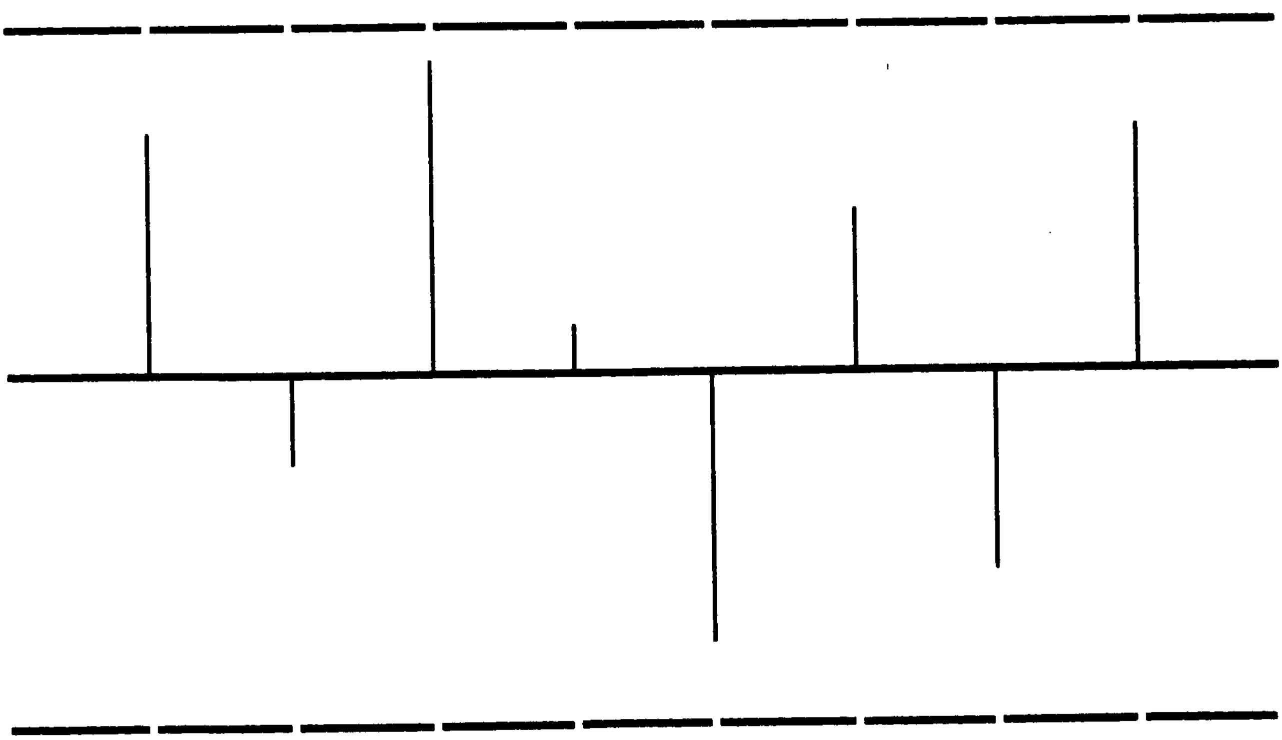


Figure 2. Example CRT Display for the Analog Visual Monitoring Task.

N88-23390

520-53
146808
11P.
S 1195835
SP

INFORMATION PROCESSING DEFICITS IN PSYCHIATRIC POPULATIONS: IMPLICATIONS FOR NORMAL WORKLOAD ASSESSMENT*

Philip D. Harvey
Departments of Psychology and Psychiatry and
Center for Cognitive and Psycholinguistic Sciences
State University of New York at Binghamton

Schizophrenic and manic patients have been described as impaired information processors since the earliest definitions of these diagnostic categories (e.g., Kraepelin, 1;2). It has taken until recent years, however, before these descriptions were developed to the point where the specific characteristics of their dysfunctions have begun to be operationalized effectively. Recent reports focusing on auditory information processing have identified several specific aspects of information processing in manics and schizophrenics that differentiate them from normals and provide ideas about group-specific aspects of performance. The characteristics of these deficits suggest in large part that psychotic information processors perform in certain ways that could be seen to be qualitatively similar to normals, but operating at lower levels of performance and being more responsive to overloading conditions.

For example, Oltmanns (3) found that both manics and schizophrenics were more distractible than normals in processing both digits and words in the presence of similar distracting information. In a closer examination of the word-span task, he found that the distraction deficits of the schizophrenics were specific to the primacy portion of the serial position curve of the presented information. He also found that schizophrenics did not shift effort to process irrelevant information, but were apparently impaired in the processing of relevant information in the presence of irrelevant information. His interpretation was that distraction impaired schizophrenics' ability to process information when higher-level cognitive processes were required, but that their processing deficits were not qualitatively different from an overloaded normal processor.

In a similar study, Pogue-Geile and Oltmanns (4) used a dichotic shadowing task to examine distraction effects in schizophrenics, manics, depressives, and normals. They found that none of the subject samples was affected by being required to shadow information in the presence of an irrelevant text passage. Interestingly, the schizophrenic subjects manifested a deficit in their ability to answer content-based questions about the shadowed information presented in the presence of distraction. These results also suggest that distraction in

*This research was supported by grant number MH38431 from the National Institute of Mental Health.

schizophrenic populations interferes with higher level processes, particularly those relevant to the encoding of information for later recall.

The same general conclusions have held up across a number of studies (many of which were reviewed by Koh, 5; Neale & Oltmanns, 6; & Callaway and Naghdi, 7) of the information processing competence of schizophrenic subjects. In many different studies schizophrenics manifest deficits in tasks measuring what Schneider and Schiffrin (8) would call controlled, but not automatic, information processing. As controlled processes are defined as those that are capacity-limited and load sensitive in normals, the conclusion would appear to be that schizophrenic subjects under load simply perform like normals under a higher level of load.

The two present studies were designed to examine overload processes in schizophrenics with an eye toward several critical questions not addressed by other studies. In most earlier information processing studies, load was not manipulated directly and its effect measured. In our study number 1 we manipulated information processing load in digit serial recall and examined the overall and serial position effects. We wanted to examine the extent to which varied aspects of information processing were load responsive and exactly how much more impaired the schizophrenics were than normals at similar load levels.

The second study examined dichotic shadowing and recall of textual material that varied in terms of its organization. We examined varied aspects of both the shadowing and recall of the material, including level of organization shadowed, number of concepts shadowed, as well as more standard indices of shadowing such as percentage correctly shadowed and errors of commission. We used the same measures for shadowing and recall in order to see directly if deficits in specific aspects of shadowing (e.g., level of organization) led to recall deficits at the same level of processing. Finally, we were interested in the specific effect of distraction in order to localize its effect in terms of which aspect of performance was maximally affected.

Study 1

Subjects

Subjects in this study were 20 schizophrenics, 13 manics (bipolars), and 10 normals. All patient subjects were acute admissions to a state psychiatric center and had been assessed with a structured rating instrument (SADS; Spitzer et al., 9) and diagnosed with DSM-III (10). All normals had been screened for a personal or familial history of psychiatric care or hospitalization. All patients were examined within 10 days of their admission to treatment and the normals were matched to them on age, sex, and other demographic characteristics.

Task and Procedure

The recall task involved the presentation at a 2-second rate of digit stimuli in trial lengths of 4,6,8, or 10 digits. Four trials per length were used and the information was presented in a tape-recorded format in a fixed, random order. Subjects were given ordered recall instructions and were asked for an immediate recall of the information at the end of the trial. Subjects were not informed before the onset of the trial as to how many digits were to be presented. The undergraduate research assistant who tested the subjects stopped the tape between trials and recorded the subjects' reponses verbatim.

Results

We scored the subjects' recall protocols using free recall methods in order to avoid as much as possible modifications of the serial position curve noted by Drewnowski and Murdoch (11). We performed analyses of both total score performance and of serial position performance. The data for the total scores are presented in Table 1 and the serial position curves are presented in Figure 1.

For the total score analyses we performed a 3(Diagnosis) x 4(Trial Length) repeated-measures ANOVA, with the final factor repeated. We found a significant 2-way interaction of Diagnosis x Trial length, $\underline{F}$ (6,120)=2.92, $\underline{p}$ <.05. In order to examine this interaction, simple-effects tests were used, finding significant diagnostic effects at lengths 8 and 10 only. In both cases, Newman-Keuls Tests indicated that normals performed better than manics, who performed better than schizophrenics.

For the serial position analyses we performed Diagnosis x Position ANOVAs within each trial length. No significant effects were detected at length 4, so that length is not further discussed. At length 6, a significant effect of diagnosis was detected, $\underline{F}$ (2,37)=4.56, $\underline{p}$ <.05, with Newman-Keuls tests finding that normals performed better than manics who in turn performed better than schizophrenics. At lengths 8 and 10 significant 2-way interactions of Diagnosis x Position were detected. In order to interpret these interactions, we used Newman-Keuls tests, comparing the three diagnostic groups across the varied positions, with the results of these analyses presented in Table 2.

The schizophrenic subjects were always the most deviant on the primacy portion of the serial position curve and were never more deviant than the manics on the recency.

Discussion

On this task it appears as if schizophrenics' total performance is much like that of a normal processor under a higher load level. For example the total performance of the

schizophrenics at length 4 is similar to that of the normals at length 8 and the normals' performance at length 10 is similar to the schizophrenics' at length 6. The manics' performance was intermediary to that of the schizophrenics and normals. In the serial position analyses, particularly at lengths 8 and 10, the schizophrenics were particularly more deviant on the primacy than the other subjects, with recency performance apparently reflecting a generalized psychotic deficit. The serial position performance of the patients was particularly distorted at length 10, with both manics and schizophrenics manifesting serial position performance that was particularly poor in the recency, probably reflecting either retrieval interference effects or generalized inability to handle both item and order information in such high loads.

A general conclusion is that schizophrenics appear to function like more highly loaded normals, with primacy performance being particularly poor. Schizophrenics appear to be almost completely overloaded at length 10, with free recall scoring producing only a 42% level of performance with no recall delay or interspersed information. Relative changes in primacy performance were considerably greater for the schizophrenics than for the normals, suggesting a particular vulnerability of resource limited functions in this population.

Study 2

Subjects

Subjects in the second study were 20 schizophrenics, 16 manics, and 16 normals. The subjects were selected and diagnosed as described above and the samples of subjects in the two studies were completely independent.

Experimental Task and Procedure

Subjects were asked to shadow and recall verbatim 8 descriptive text passages. Four passages were random collections of stories about a commonplace topic (e.g., summer) and four passages were completely organized stories. The level of organization was determined to be the maximum possible according to the Waters and Lomenick (12) descriptive passage rating scale. Four stories (2 per organization level) were presented by themselves and four were presented concurrently to the presentation of distraction story read in a female voice in the unattended ear. The ear of presentation was varied across the stories in order that each subject received one target story per organization level per distraction condition per ear. Subjects were instructed to shadow the story exactly as presented and to be prepared to recall it verbatim immediately after shadowing.

Subjects' shadowing and recall were tape-recorded and were transcribed for examination. The shadowing dependent variables that were scored by raters who were blind to all aspects of the

procedure were the percentages correctly shadowed, the number of concepts (subjects of clauses) shadowed, level of organization shadowed, accurate paraphrase errors, and semantically relevant errors. Recall DV's were the number of words used in recall, the level of organization present in recall, and the number of concepts recalled.

Results

The data regarding shadowing performance are presented in Table 3 and the data regarding recall are presented in Table 4. As we are primarily interested in distraction effects and their implications for overload, the data regarding shadowing errors are not presented since no distraction effects were found to be present in the error variables for any subjects. Analyses that yielded effects other than distraction or interactions involving distraction will not be discussed either.

A significant Diagnosis x Distraction interaction was discovered for the percentage of words correctly shadowed, F $(2,49)=4.25$, $p<.05$. Simple effects tests found that schizophrenics and no other subjects were signficantly affected by the addition of distraction. For the number of concepts correctly shadowed, another Diagnosis x Distraction interaction was detected, F $(2,49)=4.29$, $p<.05$. The same pattern of group differences was found with simple effects tests: schizophrenics were the only distractible group. For the level of organization shadowed, a triple interaction of Diagnosis x Distraction x Organization was detected. Simple effects tests revealed that for both normals and manics a significant effect of organization was present and that there were no distraction effects. For schizophrenics, a different pattern of results emerged. Schizophrenics were not affected by distraction in the random passages, probably because of floor effects, but there was a sigificant reduction in the amount of organization present in organized passages in distraction relative to nondistraction.

For the recall variables, the only variable that produced an interaction involving distraction and diagnosis was the level of organization at recall. That variable generated a significant triple interaction of Diagnosis x Distraction x Ear, F $(2,49)=3.20$, $p<.05$. Simple effects tests were used to interpret the interaction. Schizophrenic subjects had the most interesting results, where it was discovered that they manifested a right ear advantage for recall of structural information of organized passages under distraction and a left ear advantage for recall of structure of organized passages under nondistraction conditions.

Interestingly, in none of the groups was any of the shadowing and recall variables correlated, suggesting that they are measuring largely unrelated aspects of recall performance. Furthermore, within all subject groups, all the shadowing variables and all of the recall variables are correlated with each other.

Discussion

In this study we have found that distraction has a relatively specific effect of cognitive processing in schizophrenia. It appears as if distraction disrupts the ability to effectively shadow information to a greater extent than it disrupts the ability to encode information for recall. It is possible, of course, since distraction did not completely disrupt shadowing for schizophrenics, that the distraction manipulation was simply not powerful enough to interfere with encoding performance. It may be that the act of shadowing serves to focus attention to the extent that encoding can be accomplished despite any interference provided by the presence of distracting information. In addition, manic subjects performed essentially the same as normals, not being affected by distraction to any significant extent and manifesting relatively normal recall of the information presented.

Our results clearly suggest that overload effects in schizophrenics need to be carefully examined and that assumptions about the relative similarity between tasks may need to be tested. Obviously the processes of encoding for recall have some commonalities with the processes that are operating during the shadowing process. It seems, however, as if the moment-to-moment monitoring processes involved in shadowing are either more disruptive than the processes involved in encoding or that they are responsive to lower levels of interfering information.

General Discussion

If one allows the assumption that our first study has demonstrated that schizophrenics perform similarly to more highly loaded normals, then the results of the two tasks have expanded our knowledge of what might happen to normal operators during overload in shadowing. It might be the case that shadowing problems due to overload would not be reflective of the actual extent to which an operator has processed a message. Even if the basic organizational structure of the passage is appreciably disrupted, as happened to our schizophrenic subjects in the shadowing study, the extent to which the message is recalled is not impaired. This finding holds up with multiple indices of recall, including verbatim, gist, and structure aspects. One should expect, then, that normal operators who are called upon to monitor a message and then to recall or use the information from it may perform substantially better at the recall task than the shadowing task, even under high load demands. This finding would be expected even if the operator was instructed that the two tasks had equal performance priority. It might be hypothesizd that if the recall task was given higher priority than the monitoring/shadowing task that this performance discrepancy under load would be even more greatly enhanced. Whether the reverse would be true and if shadowing could be more highly prioritized than encoding is an empirical question.

It is possible that the reason that disrupted shadowing performance failed to predict recall failures is that the two processes operate completely independently of each other. A more plausible notion is that the two operate from a common resource pool with differential demands on central processing capacity. Recall that subjects were instructed to both shadow and encode for recall simultaneously and that only one of these two simultaneous processes was disrupted in the schizophrenic patients. It is possible that shadowing is more resource demanding than encoding and as a result this task was more affected by the effort involved in ignoring the irrevelant distractor story. It could also be that prioritization processes themselves are affected by distraction in schizophrenics, so that they could not effectively split their effort and perform two simultaneous processes without problems. It turned out that all subjects were better at shadowing random than organized passages and that all subjects were better at recalling organized passages. Conceivably the optimal level of textual coherence differs depending on whether text is to be recalled or only shadowed. Possibly shadowing is most effectively done on a sentence by sentence basis, with higher level organization information leading only to interference with the process. In contrast, the presence of higher level organizational features has already been demonstrated to enhance the process of recalling textual information. Viewing shadowing and recall tasks as a dual-task method may be the most productive way to further clarify the state of knowledge in this area.

Across these two studies, however, we have seen that schizophrenic information processors do not differ qualitatively from normals. We have also seen that it may be possible to draw inferences about high-level overload in normals by comparison of their performance with those of a population of subjects whose information-processing capabilities are qualitatively similar to normals but impaired in certain capacity-related ways. The use of other information-processing impaired populations may be an effective modality to generate hypotheses about abnormal or special mental states in normal subjects.

References

1. Kraepelin, E. (1919). Dementia Praecox and Paraphrenia. Livingstone, Edinburgh.

2. Kraepelin, E. (1921). Mania depressive insanity and Paranoia. Livingstone, Edinburgh.

3. Oltmanns, T.F. (1978). Selective attention in schizophrenic and manic psychoses: The effect of distraction on information processing. Journal of Abnormal Psychology, 87, 212-225.

4. Pogue-Geile, M.F. & Oltmanns, T.F. (1980). Sentence perception and distractibility in schizophrenia, manic, and depressed patients. Journal of Abnormal Psychology, 89, 115-124.

5. Koh, S.D. (1978). Remembering of verbal materials by schizophrenic young adults. In. S. Schwartz (Ed.). Language and cognition in schizophrenia. HIllsdale, NJ, Erlbaum.

6. Neale, J.M. & Oltmanns, T.F. (1980). Schizophrenia. New York: Wiley.

7. Callaway, E., & Naghdi, S. (1982). An information processing model for schizophrenia. Archives of General Psychiatry. 39, 339-347.

8. Schneider, W., & Schiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. Psychological Review, 84, 1-66.

9. Spitzer, R.L., Endicott, J., & Robins, L. (1978). The schedule for affective disorders and schizophrenia. New York, Biometrics Research.

10. American Psychiatric Association. (1980). Diagnostic and statistical manual of mental disorders, 3ed. Washington, DC.

11. Drewnowski, A., & Murdoch, B.B., Jr. (1980). The role of auditory features in memory span for words. Journal of Experimental Psychology: Human Learning and Memory, 6, 319-332.

12. Waters, H.S., & Lomenick, T. (1983). Levels of organization in descriptive passages: Production, comprehension, and recall. Journal of Experimental Child Psychology, 35, 391-408.

Table 1

Total Performance in the Digit Span Task

Trial Length	Group: Schizophrenic M	SD	Manic M	SD	Normal M	SD
4	.83	.27	.92	.17	1.00	.00
6	.65	.23	.82	.12	.93	.05
8	.49	.19	.67	.20	.85	.05
10	.42	.17	.52	.19	.78	.09

Table 2

Between Group Differences in Serial Position Performance[a]

Serial Position	Length 8	Length 10
1	n=m>s	n>m>s
2	n=m>s	n>m>s
3	n>m>s	n>m=s
4	n>m>s	n>m>s
5	n>m>s	n>m=s
6	n>m=s	n>m=s
7	n>m=s	n>m=s
8	n=m=s	n>m=s
9	---	n>m=s
10	---	n>m=s

[a] n = normal
m = manic
s = schizophrenic

Table 3

Shadowing Performance and Error Measures

		Schizophrenics								Manics								Normals							
	Ear	Organized				Random				Organized				Random				Organized				Random			
		ND		D		ND		D		ND		D		ND		D		ND		D		ND		D	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Percent Correct	R	86.6	10.95	75.05	27.95	89.4	20.65	79.3	24.8	85.88	19,88	80.31	21.1	86.06	18.04	80.5	18.74	86.5	24.67	84.50	25.57	89.00	23.8	87.5	24.6
	L	87.6	20.18	73.25	24.27	84.55	24.98	79.35	22.57	81.19	19.74	76.44	22.7	81.44	22.92	76.38	23.71	86.13	29.07	83.94	26.24	87.75	26.41	84.56	27.63
Number of Concepts	R	8.90	2.25	8.20	2.65	9.15	2.23	8.50	2.63	8.75	1.95	9.13	1.45	9.00	1.26	8.63	1.71	9.00	2.28	8.81	2.23	9.00	2.22	8.94	2.38
	L	8.90	2.22	8.35	2.32	8.80	2.44	8.15	2.37	8.44	1.90	8.56	1.93	8.44	2.28	8.38	1.89	8.94	2.59	8.63	3.03	8.94	2.62	8.81	2.54
Level of Organization	R	6.35	1.57	5.20	2.21	1.00	0	1.00	0	6.19	1.64	6.06	1.84	1.00	0	1.00	0	6.25	1.88	6.13	2.03	1.00	0	1.00	0
	L	6.45	1.47	5.45	1.93	1.00	0	1.00	0	6.06	1.24	5.44	2.10	1.00	0	1.00	0	6.19	2.04	6.19	2.04	1.00	0	1.00	0

Table 4

Recall Performance Measures

		Schizophrenics								Manics								Normals							
	Ear	Organized				Random				Organized				Random				Organized				Random			
		ND		D		ND		D		ND		D		ND		D		ND		D		ND		D	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Number of Concepts	R	3.80	2.40	4.35	1.87	2.70	1.59	2.65	2.03	4.00	2.53	4.56	2.68	3.31	2.27	3.31	1.92	6.00	1.97	7.00	1.71	5.06	1.24	4.13	1.67
	L	3.80	2.48	4.40	2.33	2.90	1.62	2.45	1.64	4.44	1.82	4.63	2.87	3.00	1.71	2.50	1.10	5.75	2.14	6.69	1.45	5.00	1.83	4.63	1.82
Words Used	R	45.05	22.23	50.50	23.33	41.75	18.96	46.10	27.59	73.19	37.83	73.63	48.74	69.31	38.70	69.25	41.41	63.94	16.85	66.13	19.53	56.13	17.93	57.88	17.78
	L	45.95	23.74	48.85	28.59	40.10	22.96	41.55	22.45	76.19	53.83	73.50	41.56	75.13	56.58	75.31	68.24	60.75	23.46	68.44	16.44	59.56	19.51	58.00	22.89
Level of Organization	R	2.55	1.32	3.05	1.90	1.20	.62	1.40	.99	3.19	1.64	3.00	1.63	1.50	.89	1.13	.34	4.13	2.36	4.44	1.79	1.00	0	1.19	.54
	L	3.30	2.18	2.60	1.47	1.20	.62	1.10	.31	2.63	1.67	2.75	1.84	1.56	1.55	1.56	1.21	3.56	2.03	5.00	2.03	1.00	0	1.50	1.32

ORIGINAL PAGE IS
OF POOR QUALITY

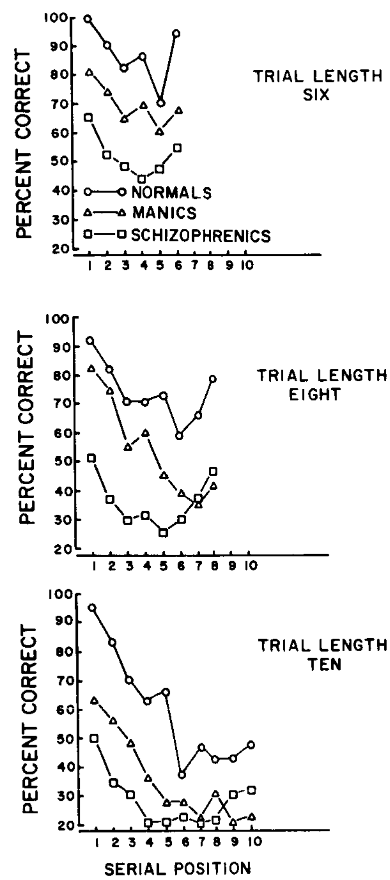


Figure 1. Serial Recall Performance Across the Varied Trial Lengths

N88-23391

S1-53
140809
118

NEUROPHYSIOLOGICAL PREDICTORS OF QUALITY OF PERFORMANCE

Alan S. Gevins
EEG Systems Laboratory
1855 Folsom St.
San Francisco, CA

EG 885150

ABSTRACT

New signal processing technologies have been developed to measure spatiotemporal neurocognitive processes of the human brain. In one experiment, application of these technologies produced measurements of distributed preparatory sets which predicted the accuracy of subsequent performance. In another experiment, neuroelectric changes were found in Air Force test pilots during the incipient stages of fatigue before behavior had severely degraded.

THE METHOD OF EVENT-RELATED COVARIANCES (ERCs)

Overview.

We have been developing new methods for recording and analyzing task-related, spatiotemporal neurocognitive patterns from the unrelated electrical activity of the brain (refs. 1-14). Since neurocognitive processes are complex, we are concerned with spatiotemporal task-related activity recorded by many (currently up to 64) scalp electrodes in many (currently up to about 25) time intervals spanning a 4-6 second period extending from before a cue, through stimulus and response, to presentation of feedback about performance accuracy. Since goal-directed behaviors require integrated processing among many brain regions, we developed the method of event-related covariance (ERC) to measure salient aspects of the brain's distributed processing networks.

The basis for ERC analysis lies with prior animal studies that have shown that when a brain region becomes involved in task performance, synchronization of a subset of neurons in that region is manifested as a change in the waveshape of its extra-cellularly recorded low frequency macropotentials (review in ref. 8). Since waveshape similarity and timing of macropotentials from different areas of the brain can be measured by covariance and correlation, these measures may characterize the spatial organization of coordinated functional activity of the areas involved in a goal-directed behavior.

PRECEDING PAGE BLANK NOT FILMED

Computing ERCs.

A number of steps are currently performed in computing ERCs. The first pass reduces spatial smearing and then selects intervals and trials with task-related information to enhance the signal-to-noise ratio and reduce the amount of data prior to measuring ERCs. The second pass measures ERCs on band-pass-filtered, enhanced averages from the reduced data set.

The steps include: 1) recording at least 50-100 trials of each task using at least 24 electrodes; 2) removing the effect of the reference channel and reducing spatial blur; 3) removing data with artifact contamination; 4) finding trials with consistent event-related signals and computing enhanced averages; 5) selecting digital bandpass filters and intervals for measurement by examining ERPS, amplitude distribution maps and Wigner Distributions; 6) computing multilag crosscovariance functions between all pairwise channel combinations of the enhanced averages in each selected analysis interval; 7) using the magnitude of the maximum crosscovariance function and its lag time as features characterizing the ERC; 8) estimating significance of ERCs by the standard deviation of the "noise" ERC; 9) graphing the most significant ERCs in each interval; and 10) statistically comparing ERC maps between conditions.

The results of ERC analysis are expressed as color graphs. Since color photographs are not possible in these proceedings, the interested reader is referred to the published literature cited in this paper.

Validation of ERCs.

ERC analysis has been applied to data recorded from several experiments. The validity of the method is demonstrated in analyses of visual stimulus processing and response execution intervals of a visuomotor task (refs. 5; 13). As predicted by neuroanatomical theory and clinical neuropsychological studies, ERC patterns corresponding to the visual stimulus processing interval involved posterior sites that led anterior parietal sites and premotor sites (Fig. 1).

While ERC patterns appear to reflect the functional coordination of immediately underlying cortical areas, we must emphasize, however, that the actual neural sources of the ERC patterns are, in fact, not yet completely known. Determining the distributed source network producing the scalp ERC patterns is the major focus of our current technical efforts.

APPLICATION TO PREPARATION AND PREDICTING PERFORMANCE

Procedure (refs. 5; 13).

Seven healthy, right-handed male adults participated in this study. A visual cue, slanted to the right or to the left, indicated to subjects to prepare to make a response pressure with the right or left index finger. One second later, the cue was followed by a visual numeric stimulus (number 1-9) indicating that a pressure of .1 to .9 kg should be made with the index finger of the hand indicated by the cue. Feedback indicating the exact response pressure produced was presented as a two-digit number one second after the peak of the response pressure. On a random 20% of the trials, the stimulus number was slanted opposite to that of the cue, and subjects were to withhold their responses on these "catch trials". The next trial followed 1 sec after disappearance of the feedback. Subjects each performed several hundred trials, with rest breaks as needed.

Twenty-six channels of EEG data, as well as vertical and horizontal eye-movements and flexor digitori muscle activity from both arms, were recorded. All single-trial EEG data were screened for eye-movement, muscle potential and other artifacts. Contaminated data were discarded.

Intervals used for ERC analysis were centered on major event-related potential (ERP) peaks. ERCS were computed between each of the 120 pairwise combinations of the 16 nonperipheral channels in intervals from 500 msec before cue to 500 msec after the feedback.

Data sets were separated into trials in which subsequent performance was either accurate or inaccurate. Accurate and inaccurate performance trials were those in which the error (deviation from required finger pressure) was less than or greater than, respectively, the mean error over the recording session.

Results and Discussion.

ERC patterns during a 375-msec interval centered 687 msec post-cue (spanning the late Contingent Negative Variation; CNV) involved left prefrontal sites, regardless of subsequent accuracy, as well as appropriately lateralized central and parietal sites (Fig. 2). Inaccurate performance by the right hand was preceded by a highly simplified pattern, while inaccurate performance by the left hand was preceded by a complex, spatially diffuse pattern.

When the trials of each of the 7 subjects were classified by equations developed on the trials of the other 6 subjects, the overall discrimination was 59% ($p<0.01$) for right hand and 57%

(p<0.01) for left-hand performance. For the subject with the most trials, average classification of 68% ($p<.001$) for subsequent right- and 62% ($p<.01$) for subsequent left-hand performance was achieved by testing a separate equation on each fifth of his trials, formed from the other four fifths.

An ERC pattern involving covariances from midline parietal, left parietal, midline anterocentral and right frontal and anterocentral sites was common to feedback to both accurate and inaccurate right- and left-hand responses. When responses were inaccurate, however, the feedback pattern additionally included the midline and left-frontal sites.

We suggest that our pre-stimulus ERC patterns characterize a distributed preparatory neural set related to the accuracy of subsequent task performance. This set appears to involve distinctive cognitive (frontal), integrative-motor and lateralized somesthetic-motor components. The involvement of the left-frontal site is consistent with clinical findings that preparatory sets are synthesized and integrated in prefrontal cortical areas, and with experimental and clinical evidence indicating involvement of the left dorsolateral prefrontal cortex in delayed response tasks. A midline anterocentral integrative motor component is consistent with known involvement of premotor and supplementary motor areas in initiating motor responses. The finding of an appropriately lateralized central and parietal component is consistent with evidence from primates and humans for neuronal firing in motor and somatosensory cortices prior to motor responses.

We further speculate that involvement of the midline anterocentral site following feedback to both accurate and inaccurate performance may reflect "motor recalibration" consequent to feedback information. Feedback-specific "updating" may be reflected by the involvement of the right prefrontal site for both accurate and inaccurate performance; behavioral verification, given feedback about inaccurate performance, by the left prefrontal site.

APPLICATION TO MEASURING EFFECTS OF INCIPIENT FATIGUE

Procedure (ref.15).

After learning and practicing a battery of tasks until their performance was stable on one day, each of five U.S. Air Force test pilots returned to the laboratory the next morning and performed the tasks for about 6 hours. Following a dinner break, they resumed task performance for an additional 6 to 8 hours.

There were four tasks in the battery, including easy and difficult continuous and discrete visuomotor tracking tasks, a simple numeric memory task, and a difficult visuomotor memory

task (VMMT). Since we expected that early neural signs of fatigue would be most evident during demanding tasks, we analyzed the VMMT first. This task required subjects to remember two continuously changing numbers, in the presence of numeric distractors, in order to produce precise finger pressures. Each trial consisted of a warning symbol followed by a single-digit visual stimulus to be remembered, followed by the subject's finger-pressure response to the stimulus number presented two trials ago, followed by a 2-digit feedback number indicating the accuracy of the response. For example, if the stimulus numbers in five successive trials were 8, 6, 1, 9, 4, the correct response would be a pressure of .8 kg when seeing the 1, .6 kg for the 9, and .1 kg for the 4. To increase the task difficulty, subjects were required to withhold their response on a random 20% of the trials. These "no-response catch trials" were trials in which the current stimulus number was identical to the stimulus two trials ago.

Trials early in the recording session with accurate finger pressures formed the "Alert" data set. Trials from early in the evening, when performance was just starting to decline, formed the "Incipient Fatigue" data set. For each subject, trials with relatively inaccurate responses were then deleted from the Incipient Fatigue data set so that the final Alert and Incipient Fatigue data sets consisted of trials with equivalently accurate performance. This crucial step allowed measurement of neuroelectric patterns associated with incipient fatigue while controlling for those due to variations in performance accuracy.

EEGs were recorded with either 33 or 51 channels with a nylon mesh cap. Vertical and horizontal eye movements were also recorded, as were the responding flexor digitori muscle potentials, electrocardiogram and respiration. Three-axis Magnetic Resonance Image scans were made of 3 of the 5 subjects.

Grand-average (over the five pilots) event-related potentials (ERPs) were time-locked to presentation of the numeric stimulus. Incipient-Fatigue ERPs were subtracted from Alert ERPs in order to highlight changes due to fatigue. Spatiotemporal neuroelectric patterns were then quantified by measuring ERCs between all 153 pairwise combinations of the 18 nonperipheral electrodes. ERCs were measured across brief segments of grand-average Alert-minus-Incipient-Fatigue subtraction ERPs. The first ERC interval was 500 msec wide and was centered 312 msec before the numeric stimulus. The next two ERC intervals were 187 msec wide and were positioned with respect to the N125 and P380 ERP peaks elicited by the numeric stimulus.

Results and Discussion.

A number of significant Alert-minus-Incipient-Fatigue ERCs were found during the 500-msec prestimulus interval. Midline central, left parietal, left anteroparietal, right anterior parietal and right posterior parietal electrodes were the major ERC foci. There were no significant ERCs in the interval centered at 62 msec post-stimulus. The ERCs computed over the P380 no-response difference ERP were focused on the midline anterocentral, and right anterior and posterior parietal electrodes.

Since ERCs are signs of functional interrelationships between brain areas, the ERC changes with Incipient Fatigue suggest that dynamic functional neural networks associated with specific cognitive functions are selectively affected during early fatigue. During the prestimulus interval, when subjects were maintaining the last two visually presented numbers in working memory and preparing for the next stimulus, ERCs decreased in number in the Incipient Fatigue condition. The lack of ERC differences between Alert and Incipient Fatigue conditions during the interval centered a 62 msec suggests that the "exogenous" stages of visual stimulus processing are relatively unaffected by early fatigue. However, during the later post-stimulus interval of trials requiring an inhibition of the response, ERCs again decreased in number with Incipient Fatigue. ERCs involving anterocentral and right parietal electrodes characterized the difference between Alert and Incipient Fatigue conditions. Since precentral, central and parietal areas are implicated by neuropsychological studies in the integration of numeric, visuospatial and visuomotor processes, the subtraction ERCs suggest a change in neural systems responsible for maintaining a representation of the magnitudes of the two visually presented numbers in working memory, and for inhibiting the response based on a comparison with working memory.

Taken together, the data suggest that although neural systems responsible for primary visual stimulus processing are relatively unaffected by incipient fatigue, cortical associative areas responsible for higher cognitive functions such as working memory rehearsal, preparation, and motor inhibition are altered prior to appreciable degradations in performance.

CONCLUSIONS

The bimanual task results demonstrate that the human brain, unlike a fixed-program computer, dynamically "tunes" its distributed, specialized subsystems in anticipation of the need to process certain types of information and take certain types of action. When these preparatory sets are incomplete or incorrect, subsequent performance is likely to be inaccurate. The fact that classification of performance accuracy improved when equations were formed and tested on the same subjects suggests that single-subject equations formed from large numbers of normative

trials may make ERC patterns useful for on-line prediction of subsequent behavior.

The fatigue experiment results demonstrate the existence of "leading indicator" neuroelectric patterns which precede serious degradation of performance consequent to extended performance of a very difficult task.

These studies demonstrate the potential of new neuroelectric signal processing technologies for measuring useful predictive information about the quality of performance. With further development, it should be possible to transition these technologies from the pure research enviornment of the laboratory to application in flight simulators, and eventually in cockpits.

ACKNOWLEDGEMENTS

The study of fatigue emerged from the collaboration of researchers from: (1) The USAF School of Aerospace Medicine, San Antonio (James Miller); (2) Systems Technology Inc., Hawthorne, California (Henry Jex and James Smith); (3) Washington University, St. Louis (John Stern); and (4) The EEG Systems Laboratory.

This research was supported by The U.S. Air Force Office of Scientific Research, The U.S. Air Force School of Aerospace Medicine, The National Institutes of Neurological and Communicative Diseases and Strokes, and The National Science Foundation.

Thanks to Judith Gumbiner for manuscript preparation.

REFERENCES

1. Gevins, A.S., Zeitlin, G.M., Yingling, C.D., Doyle, J.C., , Dedon, M.F., Henderson, J.H., Schaffer, R.E., Roumasset, J.T. & Yeager, C.L. EEG patterns during cognitive tasks. Part 1: Methodology and analysis of complex behaviors. EEG Clin. Neurophysiol., 47:693-703, 1979.

Gevins, A.S., Zeitlin, G.M., Doyle, J.C., Schaffer, R.E. & Callaway, E. EEG patterns during cognitive tasks. Part 2: Analysis of controlled tasks. EEG Clin. Neurophysiol., 47:704-710, 1979.

Gevins, A.S., Zeitlin, G.M., Doyle, J.C., Schaffer, R., Yingling, C.D., Yeager, C.L. & Callaway, E. EEG correlates of higher cortical functions. Science, 203: 665-668, 1979.

2. Gevins, A.S., Doyle, J.C., Cutillo, B.A., Schaffer, R.E., Tannehill, R.S., Ghannam, J.H., Gilcrease, V.A. & Yeager, C.L. Electrical potentials in human brain during cognition: New method reveals dynamic patterns of correlation. Science, 213, 918-922, 1981.

3. Gevins, A.S., Schaffer, R.E., Doyle, J.C., Cutillo, B.A., Tannehill, R.L. & Bressler, S.L. Shadows of thought: Shifting lateralization of human brain electrical patterns during brief visuomotor task. Science, 220, 97-99, 1983.

4. Gevins, A.S., Doyle, J.C., Cutillo, B.A., Schaffer, R.E., Tannehill, R.S., Bressler, S.L. Neurocognitive pattern analysis of a visuomotor task: Rapidly-shifting foci of evoked correlations between electrodes. Psychophysiol., 22, 32-43, 1985.

5. Gevins, A.S., Morgan, N.H., Bressler, S.L., Cutillo, B.A., White R.M., Illes, J., Greer, D.S., Doyle, J.C. & Zeitlin, G.M. Human neuroelectric patterns predict performance accuracy. Science, 235, 580-585, 1987.

6. Gevins, A.S. Pattern recognition of human brain electrical potentials. IEEE Trans. Patt. Anal. Mach. Intell., PAMI-2(5), pp. 383-404, 1980.

7. Gevins, A.S. Analysis of the electromagnetic signals of the human brain: Milestones, obstacles and goals. IEEE Trans. Biomed. Enginr., BME-31(12): 833-850, 1984.

8. Gevins, A.S. Correlation analysis. In: A.S. Gevins & A. Remond (Eds.), Handbook of Electroencephalography and Clinical Neurophysiology, Vol. 1: Methods of Analysis of Brain Electrical and Magnetic Signals. Amsterdam, Elsevier, 1987.

9. Gevins, A.S. Statistical pattern recognition. In: A.S. Gevins & A. Remond (Eds.), Handbook of Electroencephalography and Clinical Neurophysiology, Vol. 1: Methods of Analysis of Brain Electrical and Magnetic Signals. Amsterdam, Elsevier, 1987.

10. Gevins, A.S. Recent advances in neurocognitive pattern analysis. In: E. Basar (Ed.), Dynamics of Sensory and Cognitive Processing of the Brain, Springer-Verlag, 1988, 88-102.

11. Gevins, A.S. & Cutillo, B.A. Signals of Cognition. In: F.H. Lopes da Silva, W. Storm van Leeuwen & A. Remond (Eds.), Handbook of Electroencephalography and Clinical Neurophysiology, Vol. 2: Clinical Applications of Computer Analysis of EEG and other Neurophysiological Signals. Amsterdam, Elsevier, 335-381, 1986.

12. Gevins, A.S. & Morgan, N.H. Classifier-directed signal processing in brain research. IEEE Trans. Biomed. Enginr., BME-33, 1058-1064, 1986.

13. Gevins, S.A., Bressler, S.L., Morgan, N.H., Cutillo, B.A., White, R.M., Greer, D. & Illes, J. Event-related covariances during a bimanual visuomotor task, part I: Methods and analysis of stimulus- and response-locked data. EEG Clin. Neurophysiol., 1987.

14. Gevins, A.S., Cutillo, B.A., Bressler, S.L., Morgan, N.H., White, R.M., Illes, J. & Greer, D. Event-related covariances during a bimanual visuomotor task, part II: Preparation and feedback. EEG Clin. Neurophysiol., 1987.

15. Gevins, A.S., Cutillo, B.S., Fowler-White, R.M., Illes, J. & Bressler, S.L. Neurophysiological patterns of operational fatigue: Preliminary results. Paper presented at the NATO Aerospace Medical Panel Symposium, Trondheim, Norway, 1987.

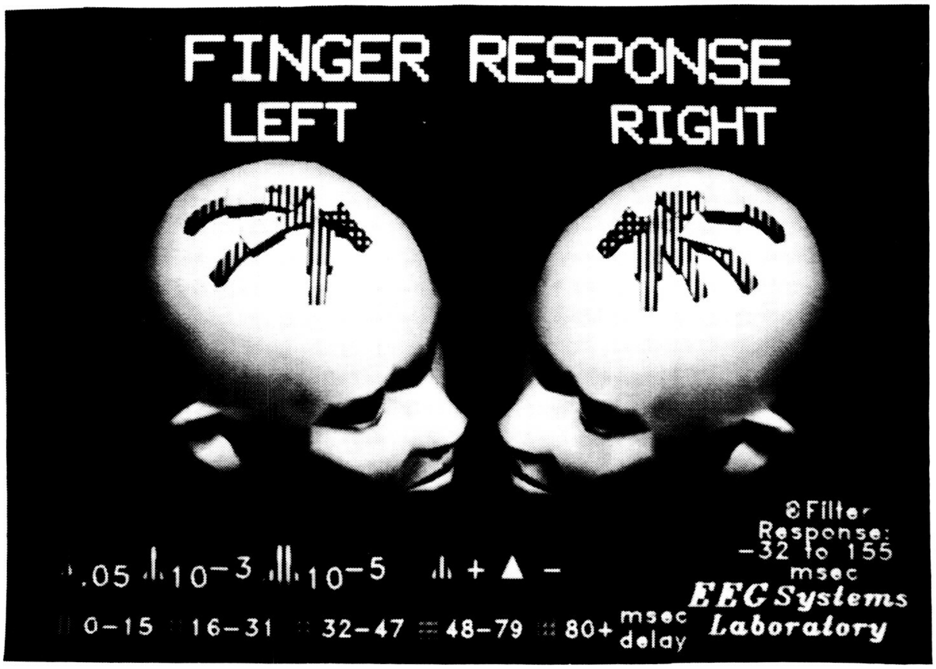
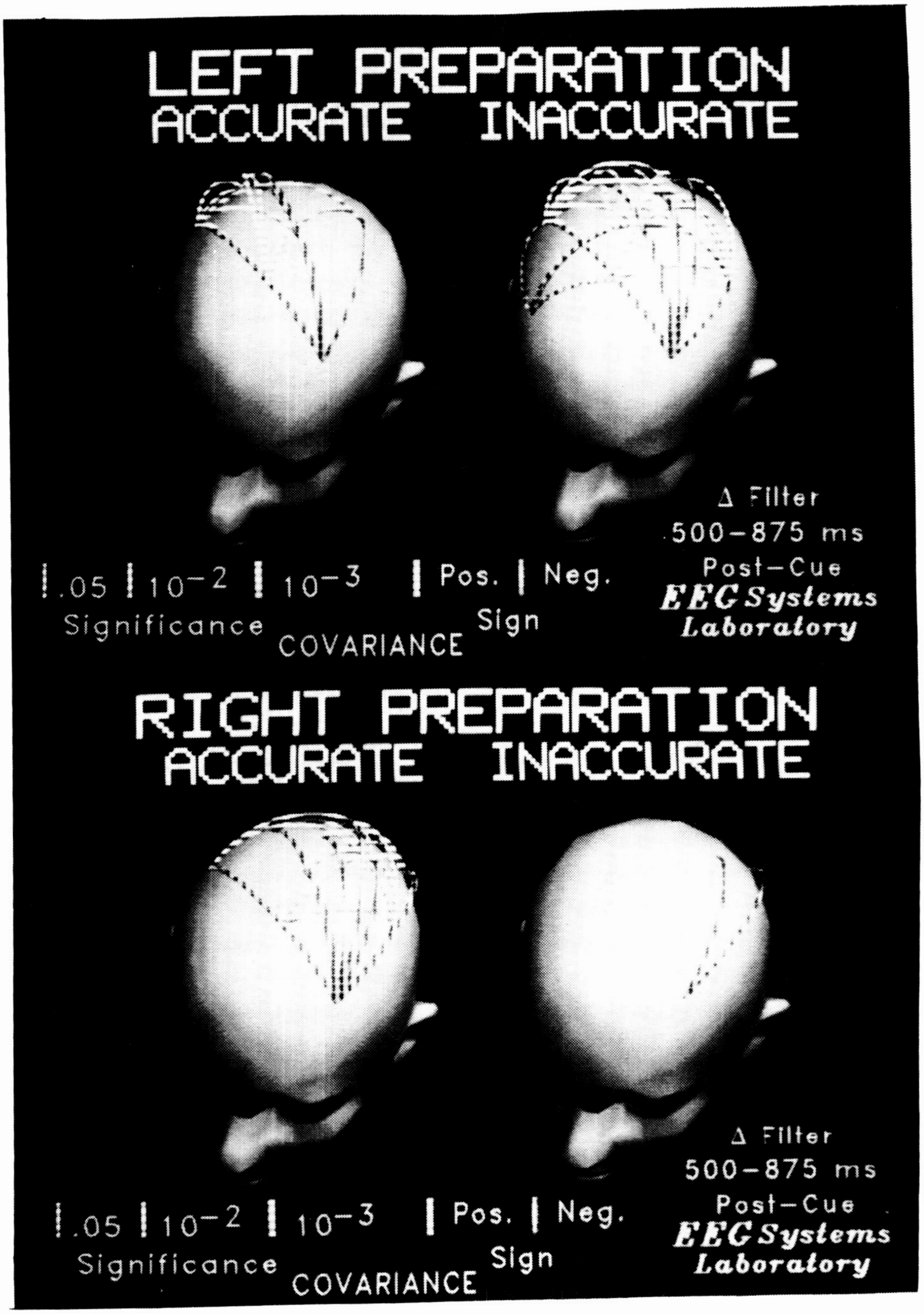


Figure 1: View of the most significant, event-related covariance patterns from the wave at the peak of a finger response. The motor-related wave was measured during a 187-msec interval centered on the peak of the left-hand and right-hand index finger pressures from theta-band filtered, seven-subject averages. The thickness of a line is proportional to tis significance (from .05 to .00005). Line pattern indicates the time delay (lag time of maximum covariance), and the arrow points from the leading to the lagging channel. ERC patterns for movement-registered timeseries also corresponded to prior functional neuroanatomical knowledge: the midline precentral electrode that overlies the premotor and supplementary motor cortices was the focus of all movement-related ERC patterns, and the other most significant ERCs involved pre- and post-central sites appropriately contralateral to the responding hand. Moreover, the pattern for the Motor Potential clearly reflected the sharply focused current sources and sinks spanning the hand areas of motor cortex.

ORIGINAL PAGE IS
OF POOR QUALITY



A

B

Figure 2: View of the most significant ($p<.05$), between-channel CNV event-related covariance patterns from an interval 500 to 875 msec after the cue for subsequently accurate and inaccurate left-hand (A) and right-hand (B) performance by seven right-handed subjects. The thickness of a line is proportional to its significance (from .05 to .005). Line pattern indicates whether covariance is positive (lighter lines) or negative (darker lines). Covariances involving left-frontal and appropriately lateralized central and parietal electrode sites are prominent in patterns for subsequently accurate performance of both hands. Magnitude and number of covariances are greater preceding subsequently inaccurate left-hand performance; fewer and weaker covariances characterize subsequently inaccurate right-hand performance.

N88-23392 522-53 140810 8P WG 032961

PHYSIOLOGICAL MEASURES AND MENTAL-STATE ASSESSMENT

John A. Stern
Washington University Behavior Research Laboratory at
Malcolm Bliss Mental Health Center
St. Louis, Missouri

I have been to a number of conferences on work load assessment in recent years. A major focus in these meetings has been on the upper end of the work load continuum. How can we evaluate, how can we reduce, how can we cope with unusually high work load levels? It is our contention that, except for relatively few situations, the real danger lies at the other end of the continuum, namely, what can we do to maintain an acceptable level of vigilance, alertness, attention, on the part of our complex equipment operator. Within this context, we immediately have to think of "mental states" (states of the organism) to make sense of work load assessment issues.

Pilots no longer fly aircraft; they exert supervisory control over a computer system which generally does a more satisfactory job of maneuvering the aircraft from take-off to landing than a human pilot. These systems, however, occasionally break down, and the pilot has to assume responsibility for flying the aircraft. How do we maintain the pilot's proficiency to fly the aircraft? How do we assure ourselves that the pilot is attentively monitoring equipment to detect and correct equipment malfunction? How do we keep him scanning the skies and his radar display to assure himself that he or someone else is not on a collision course with him/her?

We know that man's ability to monitor equipment that seldom breaks down is, at best, mediocre. What can we do to enhance vigilance? What can we do to detect or avoid vigilance decrements?

Vigilance, arousal, alertness, and attention are all concepts that touch on the issue of mental state assessment. How have psychologists traditionally gone about the task of mental state assessment? We ask subjects to rate or otherwise evaluate their state. We monitor aspects of performance and infer mental state from performance, or we can monitor physiological measures and infer mental state from the outputs of physiological sensors, or we can look at a combination of performance and physiological measures.

As a human psychophysiologist, I am interested in using physiological measures to allow me to make inferences about our subject's level of alertness, about cognitive operations used to solve problems, and about affective states. In the present context, I am interested in the impact of these mental states on performance. I am concerned with using physiological measures to predict and, hopefully, abort performance decrements or human error. I, thus, would like to have some valid measures of performance.

I am less concerned with, and interested in subjective reports, with what the subject verbalizes about either his level of alertness or ability to perform. As one trained in clinical psychology, I have little faith in what we say about issues, such as alertness and ability to perform. Human error is a major cause of all accidents. I am certain that most persons involved in an accident did not do so voluntarily. We have accidents

because our judgment about our ability to perform is in error!

We have approximately 35,000 fatal automobile accidents each year, and probably ten times that many non fatal accidents. Most of these accidents are not single, but multiple vehicle accidents. Although I object to people injuring themselves, for both humanitarian and health care cost containment issues, the courts apparently are ambivalent on this issue. On the one hand, they have declared laws insuring the wearing of safety helmets on the part of motorcyclists invalid, while on the other, they have passed laws to encourage the use of seat belts.

I object vehemently to incapacitated drivers engaging in involuntary manslaughter, or seriously injuring innocent motorists. If the law does not allow me to protect a fool from himself, it is reasonably positive about attempting to protect others from the fool (e.g., drunk driving laws). My ideal is to have each vehicle equipped with a red light that warns others when the driver is not performing safely. One can then pull off to the side of the road until the danger has passed. If the courts don't want to protect people from foolishly killing themselves, perhaps they can be encouraged to help assure some increment in safety for the innocent bystander.

How might we go about this task of evaluating mental state to reduce mayhem on the highway, and to a lesser extent, in the sky? It is our contention that as the task requirements made of drivers or pilots decrease beyond current levels, the likelihood of occurrence of accidents will increase. Although I was unable to find the documentation for the following statement, it is a reasonable one, namely, the likelihood of a driver utilizing cruise control for highway traveling increases the likelihood of his being involved in an accident. The availability of cruise control takes away a number of requirements on the driver, namely, checking his speed, varying pressure on the gas pedal to maintain a desired speed, and, to a lesser extent, checking signs indicating speed limits. The driver has to attend to a more limited set of environmental inputs. If work load falls below a given limit, we suspect that drivers may begin to reduce attention to levels where unusual environmental events may be missed--and accidents occur.

Paradoxically, rather than having more time to devote to visual scanning, steering and braking, taking away the requirement to monitor and control speed leads to a reduction in such behavior. The same situation prevails, we believe, in commercial aircraft, not of the future, but the present. BOAC pilots, as we understand it, spend most of their flights monitoring equipment, rather than being actively engaged in flying the aircraft for which they are responsible. On a minimum number of flights, they are permitted to control the aircraft during take-off, flight, and landing. We suspect that the pilot's ability to detect and correct problems, should they occur, is seriously compromised by making the pilot a monitor of displays, rather than responsible for flying the aircraft.

How can we deal with this problem? Two complementary procedures are envisioned. First, if we can monitor the pilot's level of alertness or attention to his displays, and identify periods where his attention level falls below acceptable limits, we can provide him and others with feedback

about his condition. Much like my warning signal on the top of cars that alerts other drivers that our vehicle is not being safely driven, we would like to warn the pilot, copilot, and other flight personnel when a member of the flight crew's level of alertness to his task falls below acceptable limits. Secondly, we believe such monitoring might be used to determine optimal conditions of pilot-aircraft interaction that will maintain an acceptable level of attention on the part of the flight crew. Thirdly, it could be used in the design of the cockpit of the future.

What should be monitored physiologically to evaluate alertness and attention? We do not believe that a "universal alertness monitor," which tracks physiological systems A, B, C,...N, and uses this information the same way, regardless of who the pilot is, can be designed. There are marked individual differences in physiological system responsiveness, which suggests a monitoring package unique to each individual. We will return to this issue after we explore the issue of monitoring attention. A major attentional component, for the pilot, deals with visual inputs, be they from the instrument panel or the world outside the cockpit. Auditory components, in the form of communication functions, are generally handled by the copilot; however, other auditory inputs fall in the domain of the pilot. We will single out visual input as a component that is most important to pilot function, and one that has the advantage of being able to be monitored remotely. The evaluation of attentional variables suggests that the pilot engage in definable amounts of visual scanning during most portions of the flight. Thus, fixation pause duration suggests itself as an important component. If fixation pauses exceed a specifiable upper limit (during specific flight segments), we suspect that the pilot is no longer "looking," but is "staring" (perhaps vacuously), and not taking in visual information. Pilots should check specific instruments at definable intervals. If the interval between such checks exceeds specifiable limits, we suspect the pilot is no longer flying safely. One can take this issue a step further, and evaluate patterns of instrument checks.

If dwell time on an instrument becomes unusually short, and/or the pilot returns gaze to that instrument again, shortly after having looked at it, one can again infer inefficient search. Neville Moray inferred that this pattern might suggest that the pilot acquired necessary information from an instrument, but forgot the information and had to cross-check. If this occurs "frequently," our pilot is, again, not functioning efficiently.

What other information can we obtain from eyes to monitor attention? As you might expect, I will offer the eye blink as a second variable that may provide us with useful information about visual monitoring ability. We have some information which suggests that in the performance of critical visual tasks, blinks are least likely to occur as the eyes move to the instrument that provides such information, and most likely to occur as gaze returns to a routine area of the display. Our impression, based on data collected in a DC-9 simulator at Langley, suggests that blinks are more likely to be associated with gaze shifts in the vertical, than horizontal plane, and, from other work, we know that they are more likely to be tightly coupled to large amplitude saccades and head movements. We suspect that breakdowns in attention will lead to altered patterns of saccade/blink, head movement/blink activity, and saccade/head movement activity, as well as alterations in the temporal patterning of these actions.

We can, of course, monitor eye closures and their duration. Eyelid closures in excess of .5 sec would index lack of attention to the task at hand. Thus, monitoring aspects of oculomotor activity appears, to us, to be a most reasonable procedure for evaluating changes in visual attention. We have given a few examples of what might, in general, be monitored to evaluate aspect of visual attending.

What about the issue of alertness, a necessary, but not sufficient condition for monitoring attention? We think of alertness as the readiness to respond to unusual events, while attention deals with a focus on specific events.

We, thus, need to be alert to the occurrence of unusual and infrequently occurring events. Man's ability to maintain vigilance or alertness to such events is poor. How might we monitor this ability which may change from moment to moment, as demonstrated by investigators since the 1930's (Bills, 1937 [1]; Williams et al, 1959 [2]). Behavioral measures, in other than laboratory conditions, are of little help, since, in the real world, we never know when an unexpected event is likely to occur. The research strategy recommended by us is to utilize a series of laboratory vigilance tasks and evaluate physiological measures associated with missed signals, as well as false alarms. If one can demonstrate that a given set of physiological measures are correlated with, or predictive of performance drop-out in a variety of vigilance tasks, we would be willing to recommend these measures for the evaluation of attentional attributes under conditions where we have no performance measure against which to compare our physiological measures.

We would like to briefly outline measures that have been used to measure more general and persistent states of alertness. These procedures have generally focused on what happens to such measures as a person goes to sleep.

1) Cardiac activity:

   As we move toward sleep, heart rate decreases. Whether that decrease is secondary to a decreases in motor activity, or whether it is only partially dependent or even independent of motor activity is an issue that is still being debated.

   Heart rate variability is a derivative measure, and one currently being investigated in a number of laboratories using a variety of measures of such variability. How it relates to the issue of alertness is a question that is in need of investigation.

2) Peripheral vascular activity:

   One finds a shift from vasoconstriction to dilation as the person monitored drifts toward sleep. "Spontaneous fluctuations," i.e., non-specific responses that mirror, in wave form, orienting response, might index a change in state, though it has not been systematically studied. One major problem with monitoring such activity is the sensitivity of the measure to even minor movement

artifacts.

3) <u>Skin conductance</u> (resistance or skin potential):

As a subject becomes relaxed, there is a marked decrease in skin conductance, and skin potential drifts from a large, negative value (-70 mv) toward Ø, and may even go positive.

A derivative measure of some interest here also deals with "spontaneous fluctuations." The frequency of such responses decreases as one goes toward sleep.

4) <u>Electroencephalography</u> (EEG):

The EEG has been extensively used to define stages of sleep. Unfortunately, less work has been done to evaluate levels of alertness. Two major techniques for utilizing the EEG as a research and clinical tool are in current vogue. The first evaluates alterations in ongoing electrical activity of the brain, and utilizes spectral analysis to define average activity within restricted frequency bands. The second technique evaluates changes in electrical activity produced by specific stimuli. To extract the response to such signals out of the background of ongoing EEG activity, a procedure known as signal averaging is used.

Evaluating EEG spectra associated with altered states of alertness suggests that as a person becomes drowsy, there is initial general enhancement of activity in the alpha frequency band (8-12 Hz), followed by a shift in dominant activity within this band from a higher to a lower frequency. Much of this work has been done under eyes closed conditions, and is thus, probably not directly applicable to the evaluation of attention in visually demanding environments.

A new technology is developing which graphically displays changes in electrical activity over the skull surface. This technique allows one to see dynamic changes in electrical activity during task performance. Its utilization has been hampered by the fact that no procedures for quantifying the data generated have been developed. It is, thus, a technique completely dependent on the observational skill of the user.

Evoked response technology, as applied to the measurement of alertness, has some problems. If we are interested in momentary lapses in alertness, it cannot be used in its present form, since this measure forces us to look at brain responses averaged over a number of stimuli. In general, a minimum of ten trials are necessary to extract the signal of interest out of the background noise. It may be possible to evaluate ERPs to single trials, using template matching or other procedures. If these can be successfully implemented, this objection to the use of ERPs for the evaluation of momentary alterations in alertness may be discarded.

If our concern is with slowly changing states of alertness, this technique appears to be a viable one. One can, as we have described in an earlier presentation at this meeting, evaluate ERPs to either secondary tasks that are imbedded in primary task performance or deal with ERPs to irrelevant stimuli. We would suspect that as alertness lowers, the ability of the brain to time-share information processing capability between primary and secondary or irrelevant task demands is attenuated, and that ERPs to the secondary task are altered, and that their distribution over the head might change.

5) Pupillography:

Changes in pupillary diameter occur not only as a function of changes in light intensity impinging on the eye, but also as a function of task complexity, interest in the material viewed, listened to, or tasted, affective components and states of alertness. Pupil diameter decreases as alertness is lowered. The major problem with utilizing pupillography in a visually demanding environment, is the fact that the amount of light impinging on the eye is continually changing. Since pupillary diameter changes associated with this variable are significantly larger than those associated with alertness, cognitive or affective alterations evaluating the effect of these variables will not be an easy task. Such problems can be solved, but will require major efforts.

6) Oculomotor activity:

Are components of eye movements affected by alterations in alertness? A number of investigators have suggested that alterations in saccadic eye movements occur as a function of "fatigue" or alertness. The alteration is a slowing of peak velocity or average velocity, and is best seen with relatively large amplitude saccades. As we have suggested earlier, saccade frequency may be another indicator, not only of lowering in attention, but alertness, as well. The eye blink is another component of some interest (to us). To the extent that time-on-task effects reflect alterations in alertness, we can demonstrate that in vigilance tasks, there is an increase in average blink closure duration as a function of time-on-task, as well as an increase in long closure durations (closures exceeding 200 msec). Thus, the eye can provide us with useful information, not only with respect to attentional attributes, but alertness, as well.

7) Body movements:

We know of no data dealing with the effect of alterations in alertness on body movements. We suspect that as a person drifts toward drowsiness, he may initially demonstrate increases in body movements, followed by a precipitous decline in such movements, prior to closing the eyes and drowsing off.

These are a few examples of physiological and behavioral variables that should be investigated with respect to their utility in measuring alterations in attention and alertness. We have described a number of measures, and suspect that the best measure of alertness would utilize a combination of such measures. The combination would be individualized to maximize their predictive utility. A lot of research still needs to be done before we achieve this state.

REFERENCES

1. Bills, A.G. (1931) Blocking: a new principle in mental fatigue. American Journal of Psychology, 43, 230-245.

2. Bills, A.G. (1937) Mental fatigue and anoxia compared. Journal of Experimental Psychology, 20, 437-452.

BIBLIOGRAPHY

1. Williams, H.L., Lubin, A., and Goodnow, J.D. (1959) Impaired performance with acute sleep loss. Psychological Monographs, 73, (whole #484).

N88-23393 S53-53 140811 88

A CORRELATIONAL APPROACH TO PREDICTING OPERATOR STATUS

Clark A. Shingledecker
NTI, Incorporated
Dayton, Ohio

N8385586

ABSTRACT

This paper discusses a research approach for identifying and validating candidate physiological and behavioral parameters which can be used to predict the performance capabilities of aircrew and other system operators. In this methodology, concurrent and advance correlations are computed between predictor values and criterion performance measures. Continuous performance and sleep loss are used as stressors to promote performance variation. Preliminary data are presented which suggest dependence of prediction capability on the resource allocation policy of the operator.

INTRODUCTION

Modern advances in engineering and electronics technology continue to be responsible for a phenomenal increase in the potential effectiveness of military and commercial aircraft systems. However, the enhanced speed, operating range, maneuverability, remote sensing, and weapons capabilities made possible by these technologies are also producing significant changes in the role and importance of critical flight crew members, and in the performance requirements that are imposed upon them. As a consequence, serious consideration must be given to methods and approaches which can be used to insure optimal human performance in future airborne operations.

Several factors contribute to a growing concern over the maintenance of aircrew performance. The use of increasingly sophisticated flight computers has relieved the aircrew of many labor-intensive duties, and shifted their task to one of monitoring and supervising a complex and highly flexible system. Such automation often leads to a reduction in crew size and creates a situation in which increasingly critical responsibilities are assigned to individual operators whose performance can easily become the single most important determinant of the outcome of a major battle or of the safety of hundreds of passengers.

The problem of reduced crew redundancy is compounded by a concommitant increase in mental workload. The cockpits and flight decks of contemporary aircraft are capable of providing pilots with vast amounts of data that must be processed in a timely and accurate manner if system performance is to be maintained. In many cases, the resulting perceptual and cognitive task demands can approach, and even exceed, the inherently limited information processing capacities of even the most experienced personnel.

Traditionally, human factors specialists have approached the problem of supporting pilot performance through the design of crew station interfaces to minimize information overload, and through the development of improved training technologies. While these interventions have been successful, it is unlikely that they will continue to be sufficient by themselves to insure optimal system performance in an environment where pilot task demands are increasing, and pilot performance capabilities can be degraded by a variety of physical and psychological stressors. Included among the obvious threats to aircrew performance capacities are fatigue and sleep loss in extended operations, use of prescribed or illegal drugs, and in combat aircrews, exposure to chemical, biological and nuclear threats.

Taken together, the rising criticality of the performance exhibited by key crew members, growing task demands and the incapacitating potential of operational stressors suggest that specific, interactive subsystems may be needed to guard against catastrophic failures due to human error.

One technically feasible approach that has been suggested for preventing human errors would involve monitoring the performance capabilities of the human operator. At the simplest level, such biocybernetic intervention would permit the evaluation of performance capability prior to a flight in order to select those personnel who exhibit an optimal capacity to meet mission objectives. In a more advanced application, performance capabilities could be monitored on a moment-to-moment basis during a mission. Thus, impending operator performance decrements could be detected automatically, and the information used to alert the pilot, inform command personnel or even initiate computer control of the system.

The general computer hardware, software, and sensing technology is currently available to implement biocybernetic systems capable of monitoring the performance capability of human operators. However, little is presently known about the indices of human function that could be used to accurately and reliably measure and predict performance capabilities in a non-intrusive fashion. The purpose of this paper is to present a methodological approach with preliminary data aimed at identifying behavioral and electrophysiological predictors of impending performance failure.

RESEARCH METHOD

The methodology developed for this exploratory research represents a departure from classical research techniques which are employed to investigate measures of performance capability. In such traditional studies, the goal is to assess a measure's capability to reflect the presumed impact of an intervening hypothetical construct (e.g., fatigue, chemical intoxication, boredom, disease) on the human operator. Thus, these studies

attempt to show that when an independent variable such as sleep loss or time-on-task is varied, the measure under examination behaves in a manner which is hypothesized to be functionally equivalent to a concommitant change in the intervening variable (e.g., a monotonic increase in reaction time with increasing fatigue).

While such experimental approaches are acceptable in research designed to investigate specific pscychological phenomena, they are neither warranted nor appropriate when the research goal is to identify measures which predict performance change. The purpose of the methodology demonstrated in the present study is to specify metrics that predict performance variation. This purpose dictates a more operational approach where, rather than testing a hypothesis about causal factors linking an intervening variable and performance, a relationship is sought between a predictor metric and a criterion performance index.

In the present methodology, candidate performance predictor metrics are correlated with simultaneous and temporally succeeding measures of performance on a simulated systems operation task. Within this approach, predictor measures which correlate highly with performance on the criterion or primary task of interest can be considered reliable indicators of operator performance decrement.

While human performance naturally varies within a restricted range under normal conditions, the degree of variation observable over a typical experimental session is likely to be highly constrained. Thus, in the present methodology, performance variability is induced by exposing subjects to the combined stressors of sleep loss and continuous performance. It should be noted that the intent of imposing these stressors is not to produce some predicted pattern of decrement due to fatigue or diurnal cycles of performance efficiency. Instead, the technique is simply designed to capitalize on the performance variation likely to be produced by these conditions in order to examine a broad range of within-subject performance variability.

In summary, the object of the methodology is to provide a standardized approach to evaluating candidate measures which will predict reductions in performance capability. The approach is essentially correlational and is designed to provide quantitative estimates of the capacity of physiological, behavioral or subjective metrics to predict the variability of human performance on a task of interest.

A limited experimental implementation of the methodolgy has been completed in which two subjects performed a complex time sharing task continuously for eight hours following twelve preceding hours of sleep deprivation. This task was designed to simulate a generic systems operation activity (e.g., combat aircraft operation) and contained two primary components which

were performed simultaneously with equal priority. The first of these components was a manual control task.

The control task was a single axis (vertical), unstable compensatory tracking task similar to that described by Shingledecker (ref. 1). The task required subjects to view a cursor on a monochrome video monitor, and to keep the cursor centered over a fixed target by turning a control knob.

The second component of the simulated operational task was a visual monitoring task. The monitoring task is somewhat similar to that devised by Alluisi (ref. 2) and requires subjects to view four computer generated vertical displays that are similar to tape instruments. The scale on each display consists of six hash marks, and the center of the scale is indicated by a small circle. Under nonsignal conditions, the pointers located just to the left of the scale markings on each dial move from one position to another in a random fashion. The pointer movements on each dial are totally independent of the other dials, and occur at an update rate of 5 moves/sec. At unpredictable time intervals, the pointer on one of the four dials becomes biased to either the top half or the bottom half of the scale. This signifies a signal condition to which the subject is instructed to respond by pressing the appropriate key on a four-button keypad. Signals occurred at a frequency of 4 to 5 each minute.

To perform the combined tasks, the subject sat at a work station containing two video monitors. The tracking task was displayed on a screen which was located directly in front of the subject. The monitoring task was displayed on a monitor centered above the tracking monitor and tilted approximately 20 degrees toward the subject. Viewing distance for both monitors was approximately 60cm. The tracking task was controlled by rotating a knob in the horizontal plane with the dominant hand. The monitoring task responses were recorded from four push buttons controlled by the non-dominant hand.

Five candidate predictor measures were selected to match the information processing demands of the system operation task. In order to assess general activation level factors, four frequency bands of the EEG spectrum were selected for power spectrum analysis. In addition, as general measures of alertness, eyeblink closure duration and subjective fatigue metrics were employed.

A primary aspect of the simulated systems operation task was a display monitoring activity. In order to assess such perceptual demands, the visual memory search task was selected (Sternberg, ref. 3). Finally, in order to assess the response output capabilities of the operator associated with the high manual control demands of the vehicle operation task, the Interval Production Task (IPT) was used (Michon, ref. 4).

RESULTS

Data were collected on the criterion systems operation task and on the physiological metrics in five minute intervals. The interpolated behavioral measures were collected during a break period preceding each 50 minute performance period. Advance correlations between the predictor measures and criterion performance were computed for a variety of temporal relationships. However, to permit comparisons across the behavioral and physiological measures, only advance predictor correlations for the eight performance periods are discussed here. In this case, predictive relationships were assessed by correlating mean tracking and monitoring scores for each hour with the physiological metrics obtained in the preceding hour, or with the behavioral data collected during the preceding break period.

These correlations are shown in Table 1. Although the results are based on only two subjects, a number of tentative observations can be made from these data regarding the relative predictive capacity of the candidate parameters.

A strong relationship was obtained between performance and the proportion of total EEG power in each of four measured frequency bands. As shown in Table 1., both tracking error and monitoring signal misses were associated with power in each band. The pattern of correlation across the four bands is a general shift in power, such that poorer criterion performance occurred when the relative power in the low frequency band (delta, 1-3 hz) increased and relative power in higher frequency bands decreased (4-30 hz).

Similarly positive predictive relationships were obtained for the measures of eyeblink behavior. Increases in tracking error as well as poorer signal detection were predicted by larger amplitude blinks, higher blink rates, longer descent times for the eyelid, and longer closure durations.

A more variable set of relationships was obtained between the interpolated behavioral task measures and criterion performance. In general, criterion task decrements were associated with a decrease in duration of the interval between finger taps on the IPT task and an increase in the variability of intertap intervals. Longer Sternberg memory search task reaction times were also predictive of poorer criterion performance.

Although the results summarized above are generally descriptive of the average correlations between the predictor measures and criterion performance, inspection of Table 1 reveals marked individual differences between the two pilot subjects. Specifically, for Subject 1, correlation coefficients were consistently larger for the monitoring performance index

than the tracking. In contrast, the predictor measures were more strongly associated with tracking error than monitoring misses for Subject 2.

A potential explanation for this finding is apparent in an inspection of the hourly mean performance scores that were recorded on the two elements of the simulated systems operation task. Over the eight hour testing period, Subject 1 displayed no more than a 22% variation in tracking error. In contrast, monitoring performance varied as much as 60% and declined consistently across the testing sessions. The opposite pattern of performance was apparent for Subject 2 who displayed a greater decrement in tracking performance. Since the time sharing nature of the criterion task allowed the subjects to freely allocate their attentional resources to the tracking and monitoring components, these data suggest that the subjects devoted the bulk of their diminishing capacities to different components of the criterion task.

Such an explanation is congruent with the correlational findings for the physiological and behavioral predictors. Apparently, for these metrics predictive power may be dependent on the resource allocation policy adopted by the performer. Thus, in the case of the pilot subjects, performance on the interpolated behavioral tasks anticipated the component of criterion task performance that received the least effort expenditure. In support of this interpretation, subjective fatigue ratings for Subject 1 were positively related to monitoring missed detections ($r=.92$), but unrelated to tracking errors ($r=-.02$). Likewise, for Subject 2, fatigue ratings were strongly associated with tracking error ($r=.92$), but were not significantly correlated with monitoring misses ($r=.26$).

CONCLUSIONS

The results outlined above suggest that the methodological approach described in this paper can be used to identify and select reliable indicators of impending performance degradation in aircrews and in the operators of other critical systems. In order to develop practical technologies for monitoring human performance capabilities, a focused effort will be required in which these techniques are exercised to specify useful parameters, to validate their predictive capabilities for operational situations, and to embody them in field-usable hardware.

The work reported here suggests that no single index of human function is likely to provide global performance prediction in all task environments. Thus, accurate anticipation of performance degradation will probably be achieved only by a family of technologies from which appropriate measures will be selected to match operational environments. At a minimum, such matching will be based on three groups of factors.

As suggested by multi-factor models of human performance, a primary consideration will be the information processing resource structure of the operator's task. Measures which assess the integrity of perceptual, central and response processes as well as activation level will have to be selectively applied to tasks and environments which make differential demands on these resources. In addition, as the present results indicate, task priorities will have to be assessed in order to determine the specific aspects of performance that will be predicted by monitoring parameters.

A second group of matching factors is the temporal prediction requirement of the operational scenario. The complete results of the preliminary study indicated that different metrics varied in terms of the time period for which significant predictions were obtained. Thus, it will be necessary to employ these measurement methods in a selective manner to correspond with requirements for long term predictions (e.g. how likely is it that pilot "A"'s performance will be degraded in the next five hours?) and for short term, continuous prediction (e.g., is it probable that pilot "B" will commit a catastrophic error in the next few minutes?).

Finally, selection of prediction measures will also be determined by the limits and practicalities of the operational environment. For example, the potential intrusiveness of some measures may prevent their use during high demand, continuous performance missions. However, these measures may be preferable in situations where periodic, interpolated testing is possible. Other practical selection factors might include the size and weight of the monitoring equipment, and the operator's acceptance of any necessary monitoring sensors.

REFERENCES

1. Shingledecker, C. A. _A Task Battery for Applied Human Performance Assessment Research_. Air Force Aerospace Medical Research Laboratory, AFAMRL-TR-84-071, 1984.

2. Alluisi, Earl A. Methodology in the use of synthetic tasks to assess complex performance. _Human Factors_, 1967, _9_ (4), 375-384.

3. Sternberg, S. Memory Scanning: Mental processes revealed by reaction time experiments. _American Scientist_, 1969, _57_, 421-457.

4. Michon, J. A., Tapping regularity as a measure of perceptual motor load. _Ergonomics_, 1966, _9_ (5), 401-412.

TABLE 1.

Performance Prediction Correlations

EEG Proportional Power

		Delta	Theta	Alpha	Beta
Tracking Error	S1	.14	.29	-.17	-.21
	S2	.93	-.88	-.94	-.93
Missed Signals	S1	.95	-.64	-.86	-.97
	S2	.17	-.07	-.26	-.29

EOG Eyeblink Parameters

		Interval	Amplitude	Duration	Descent
Tracking Error	S1	-.14	.22	.37	0
	S2	-.76	.10	.99	.94
Missed Signals	S1	-.84	.62	.80	.81
	S2	-.22	-.02	.04	-.20

Interpolated Behavioral Tests

		IPT Duration	IPT Variability	Sternberg RT
Tracking Error	S1	-.28	.22	.01
	S2	-.66	.69	.35
Missed Signals	S1	-.76	.53	.55
	S2	-.30	-.05	-.07

N88-23394

S24-53
140812
98.
ϕ5 473598

BRAINSTEM RESPONSE AND STATE-TRAIT VARIABLES

Kirby Gilliland*
Department of Psychology
University of Oklahoma
Norman, Oklahoma

INTRODUCTION

The purpose of this paper is to summarize a series of investigations from our personality research program that have relevance for mental state estimation. For several years, we have been conducting research at the interface between the areas of personality, human performance, and psychophysiology. Of particular concern have been those personality variables that are believed to have either a biological or perceptual basis and their relationship to human task performance and psychophysiology. These variables are among the most robust personality measures and include such dimensions as extraversion-introversion, sensation seeking, and impulsiveness. These dimensions also have the most distinct link to performance and psychophysiology. Through the course of many of these investigations two issues have emerged repeatedly: a) these personality dimensions appear to mediate mental state, and b) mental state appears to influence measures of performance or psychophysiology.

This paper will provide a selective review of some of those studies that have highlighted these issues. Of particular concern will be those studies that offer specific insight into these issues or possible mechanisms for exploring them.

SOME FUNDAMENTAL DISTINCTIONS

To better understand the influence of personality variables or mental states it is important to understand the distinction between trait and state variables. Both are theoretical in nature, and both are believed to influence behavior. Traditionally, personality *trait* variables have been viewed as relatively permanent internal dispositions. That is, traits are evidenced regularly, are internal in origin, and are enduring in their nature. *States*, on the other hand, have been viewed as characteristics that are irregular and short-lived, and are usually viewed as responses to external social or environmental factors (ref. 1). While this distinction is generally accepted, it has not had universal support (ref. 1,2,3,4,5).[1]

* The presentation of this paper was supported in part by N.T.I., Incorporated, Dayton, Ohio.

1 Allen and Potkay (ref. 2) have suggested that the distinction between traits and states is arbitrary. They argue that rather than being two separate types of dimensions, traits and states are simply ends on a continuum. Further, (ref. 3) they argue that the delineation of trait or state measures is unnecessary and that researchers should simply "...adopt a more neutral, operational approach to predicting behavior."

It is also important to establish the relationship between traditional personality states and more general mental states. Personality states have typically referred to characteristics that have parallel trait measures, for example, state and trait anxiety and state and trait arousal. The term mental states refers to a much broader range of mental phenomena including such states as confusion, disorientation, boredom, and even fatigue. In this sense, personality states could be viewed as a subgroup of the broader category of mental states. Therefore, much of our research has been an exploration of a special category of mental states and its relationship to performance and psychophysiology. The remainder of this paper will concentrate on one state-trait dimension, that of arousal.

TRAIT-STATE MEASURES OF AROUSAL, PERFORMANCE, AND PSYCHOPHYSIOLOGY

Recent interest in the biological bases of personality has centered on a group of personality dimensions that are believed to share the common underlying dimension of arousal. The most intensely researched arousal-based dimension, extraversion-introversion (ref. 6), is believed to be the result of differential ascending reticular activating system (ARAS) arousal. It is believed that introverts have higher ARAS arousal levels as compared to extraverts and seek to restrict environmental stimulation in order to maintain a more comfortable overall level of arousal. Conversely, extraverts have a lower ARAS arousal level and seek higher levels of environmental stimulation to provide a more comfortable overall level of neural activity. The extraversion-introversion dimension and construct of arousal have been so closely linked they have often been viewed as synonymous. In fact, it is not unusual to find extraversion-introversion scales being used as a trait arousal measurement instrument, or as a method to "manipulate" arousal.

Typically, studies of extraversion-introversion are cast within an arousal framework and the results of these studies are also interpreted within the context of arousal dynamics. It was during these types of investigations that we began to realize that not only were introverts and extraverts performing differently, but also they were experiencing quite different mental states. For example, during a study of simple visual reaction time before, during, and after noise stress (ref. 7), extraverts and introverts not only performed quite differently but also reported quite different mental states. In this study, groups of introverts and extraverts performed simple visual reaction time during three seven-minute periods. One group of extraverts and one group of introverts simply performed reaction time throughout the overall 21-minute period. The remaining group of introverts and the remaining group of extraverts also performed simple visual reaction time throughout the 21-minute period. However, during the second seven-minute period, both of these groups were exposed to 75dB intermittent, cafeteria-type noise.

Figure 1 shows the results from this experiment. It should be noted that introverts showed an overall faster reaction time as compared to extraverts. This finding is typically explained within the context of an arousal model, and such results are viewed as supportive of the arousal-based nature of the extraversion-introversion trait. Noise exposure caused a similar degradation in reaction time performance for both extrovert and introvert groups. What was surprising was that during the post noise period, the last

seven-minute period of reaction time, introverts exposed to noise returned to a level of RT performance not unlike that of introverts not exposed to noise. Extraverts who were exposed to noise appeared to show continued degradation in performance over that resulting from noise exposure.

It is possible to construct a number of post hoc explanations for these results based on arousal theory. What is interesting about this particular study is that there is a much simpler explanation for these results. Following the completion of the first seven-minute reaction time period, each subject filled out a post-test questionnaire. Included in this questionnaire were a number of questions regarding mental state; for example, subjects were asked to rate their level of interest, boredom, and frustration. They were also asked to rate the amount of time they performed the simple visual reaction time task. In analyzing the results of this post-test questionnaire it was learned that extraverts were significantly more bored and frustrated with the task as compared to introverts. In addition, extraverts rated the task as lasting twice as long as the introverts. Thus, introverts and extraverts appeared to experience quite different mental states during the performance of this experiment.

This study, as well as many others, have shown what appear to be important trait arousal differences between introverts and extraverts. In the present study this can be seen in the overall faster reaction times of introverts as compared to extraverts. However, this study also demonstrates that environmental variables (in this case a lack of stimulation) can differentially influence the mental states of introverts and extraverts leading to quite different performance.

In subsequent investigations[#,+] (ref. 8), we attempted to explore more directly a link between extraversion-introversion and neural activity. These studies have utilized the brainstem auditory evoked response (BAER), a sensory evoked response reflecting the activity in the auditory pathway--a neural pathway that transverses the ARAS. The BAER provides an exceptionally stable measure of neural functioning in the auditory pathway. The BAER is derived by averaging the first ten msec of multiple (1000 or more) auditory pathway evoked potentials, elicited by short-latency click or tone stimuli. This average evoked potential results in seven vertex-positive waves believed to reflect sequential neural activity at successively higher levels of the brainstem auditory pathway (ref. 9, 10). The putative generators of wave I through wave VII are the acoustic nerve, the cochlear nuclei, the superior olives, the lateral lemniscus, the inferior colliculus, the medial geniculate, and the thalamocortical radiations, respectively (ref. 11, 12). It was believed that the stability of this measure and its close neural approximation to the ARAS made it a viable possibility for exploring differences between introverts and extraverts.

[#] Andress, D.: Individual Differences in Brainstem Auditory Evoked Responses. Unpublished Master's thesis, University of Oklahoma, 1981.

[+] Bullock, W.: A Converging Measures Test of Eysenck's Biologically-based Theory of Introversion-extraversion. Unpublished doctoral dissertation, University of Oklahoma, 1984.

The results of a series of studies of the BAER in introverts and extraverts[#,+] (ref. 8, 9) can be summarized in Figure 2. Introverts have been shown consistently to have wave V latencies that are significantly faster than those of extraverts. This has been the major and most consistent finding across the studies performed in our laboratory. This finding suggests that introverts have greater neural responsivity in the area of the lateral lemniscus and inferior colliculus. It is interesting to note that this area corresponds closely to the hypothalmic region that Eysenck views as the seat of arousal differences between introverts and extraverts. Thus, these studies seem to support the view that a personality dimension based primarily on arousal differences can be demonstrated by a physiological measure.

Another avenue in our research has been the exploration of BAER differences in relation to cognitive workload, or alternatively the exploration of state arousal (ref. 13, 14). In one study, BAERs were recorded during a pretest baseline period, during three (low, moderate, high) workload sessions, and during a post-test baseline period (ref. 13). The major results of this study revealed that longer latencies were produced at wave VI for all workload conditions as compared to the pretest baseline period. The BAER differences that were observed did not systematically differentiate the workload conditions represented in this study. Nor did the post-test baseline return to the pretest baseline level. However, the BAER was shown to be sensitive to state arousal manipulation when contrasting baseline and workload conditions.

These findings were replicated in a followup study of the post-test baseline recovery period (ref. 14). Subject's BAERs were recorded during a pretest baseline period, during the same three workload conditions, and during a post-test baseline period just as in the previous experiment. In addition, BAERs were recorded at five-minute intervals for forty minutes following the workload trials. Finally, BAERs were recorded during an additional trial at the high workload level in ABAB design fashion. The results of this study are illustrated in Figure 3. These data suggest that wave VI of the BAER is affected by cognitive workload in comparison to prior resting conditions (just as in the previous study). Wave VI latency does not fully recover under passive baseline measurement until after approximately 35-40 minutes. The final BAER under high workload conditions was comparable to those obtained under the earlier workload trials. Thus, the apparent covariation of wave VI latency of the BAER with cognitive workload suggests that this measure may be a responsive index of state arousal, albeit in a discrete rather than continuous fashion.

The results of these studies of BAER activity in relationship to extraversion-introversion and cognitive workload suggest an interesting possibility. One component, wave V latency, appears to differentiate reliably the construct of arousal as a trait. Wave VI latency has been shown to differentiate reliably state alterations in arousal. Thus, it is possible that the BAER may be useful as a method for assessing neural activity related to both trait and state forms of arousal.

SUGGESTIONS FOR IMPROVED TECHNOLOGY

It is unlikely that any single dependent measure will prove sufficient to capture and portray mental states. More likely complex multivariate

procedures will be needed to more fully explain the relationship between mental states, human performance, and psychophysiology. Our past research has suggested candidate behavioral and psychophysiological measures with the potential for aiding in the exploration of this complex relationship, but multivariate measurement alone will probably be insufficient to advance our understanding.

Major new advances in our knowledge of the relationship between mental state and task performance will probably be made through research integrating current advances in cognitive science, human factors, individual differences, and psychophysiology. For example, our laboratory is currently moving toward procedures that utilize careful laboratory control of environmental and task variables combined with real time multivariate data acquisition and analysis to provide a time-series based method for exploring these types of relationships. This technique will require the recording of multiple performance measures along with selected psychophysiological and subjective ratings, and displaying these outputs in real time. Using time-series based techniques one can then explore the interrelatedness of these measures and attempt to identify those measures that may be the most efficient in predicting such critical operator factors as performance efficiency, resource recruitment ability, and performance failure.

The distinguishing characteristic of this approach is one of modeling the performance "dynamically" rather than the usual static method associated with traditional experimental methodology. By using a more dynamic procedure one can explore not only the effect of some variable during a baseline and experimental phase (as is common in experimental techniques), but also the initial reaction, the long-term recruitment or compensation ability, the additive effects of stressors or drugs, and the rate of decline in performance ability. These characteristics are sometimes lost in standard experimental formats, but are often critical elements in defining the capability of human operators. Through multivariate, time-series-based techniques and the advent of high speed/capacity, real-time computer technology, we may be able to learn more about many of the operator variables, such as mental state, that significantly influence system performance.

REFERENCES

1. Zuckerman, M.: General and Situation-specific Traits and States: New Approaches to Assessment of Anxiety and Other Constructs. In M. Zuckerman & C.D. Spielberger (Eds.) **Emotions and Anxiety: New Concepts, Methods, and Applications.** Hillsdale, N.J.: Erlbaum, 1976.

2. Zuckerman, M.: The Distinction Between Trait and State Scales is *Not* Arbitrary: Comment on Allen and Potkay's "On the Arbitrary Distinction Between Traits and States". **Journal of Personality and Social Psychology**, vol. 44, 1983, 1083-1086.

3. Allen, B. P., & Potkay, C. R.: On the Arbitrary Distinction Between States and Traits. **Journal of Personality and Social Psychology**, vol. 41, 1981, 916-928.

4. Allen, B. P., & Potkay, C. R.: Just as Arbitrary as Ever: Comments on Zuckerman's Rejoinder. **Journal of Personality and Social Psychology**, vol. 44, 1983, 1087-1089.

5. Fridhandler, B.: Conceptual Note on State, Trait, and the State-Trait Distinction. **Journal of Personality and Social Psychology**, vol. 50, 1986, 169-174.

6. Eysenck, H. J.: The **Biological Basis of Personality**. Springfield, Illinois: Charles C. Thomas, Inc., 1967.

7. Gilliland, K., & Bracey, S.: Simple Visual Reaction Time During and After Noise Stress in Extraverts and Introverts. Paper presented at the meeting of the Midwestern Psychological Association, St. Louis, Missouri, 1980.

8. Bullock, W., Gilliland, K., & Kujawski, A.: Effects of Stimulus Intensity and Repetition Rate on Brainstem Auditory Evoked Response and Recovery Rate of Introverts and Extraverts. Paper presented at the meeting of the Society of Psychophysiological Research, Minneapolis, Minnesota, 1982.

9. Bullock, W., Gilliland, K., & Dannels, S.: The Effect of Simple Vigilance and Caffeine Administration on the Brainstem Auditory Evoked Response of Introverts and Extraverts. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, Illinois, 1984.

10. Jewett, D. L., Romano, M. N., & Williston, J. S.: Human Auditory Evoked Potentials: Possible Brainstem Components Detected at the Scalp. **Science**, vol. 167, 1970, 1517-1518.

11. Jewett, D. L., & Williston, J. S.: Auditory-Evoked Far Fields Averaged from the Scalp of Humans. **Brain**, vol. 94, 1971, 681-696.

12. Stockard, J. J. & Rossiter, V. A.: Clinical Pathological Correlates of Brainstem Auditory Response Abnormalities. **Neurology**, vol. 27, 1977, 316-325.

13. Gilliland, K., Shingledecker, C., Wilson, G., & Peio, K.: Effect of Workload on the Auditory Evoked Brainstem Response. **Proceedings of the Human Factors Society**, vol. 1, 1984, 37-39.

14. Gilliland, K., & Dannels, S.: Recovery Time of the Wave VI Component of the Auditory Evoked Brainstem Response Following Cognitive Workload. Paper presented at the meeting of the Southwestern Psychological Association, Austin, Texas, 1985.

Figure 1: Simple Visual Reaction Time of Extraverts and Introverts Before, During, and After Noise

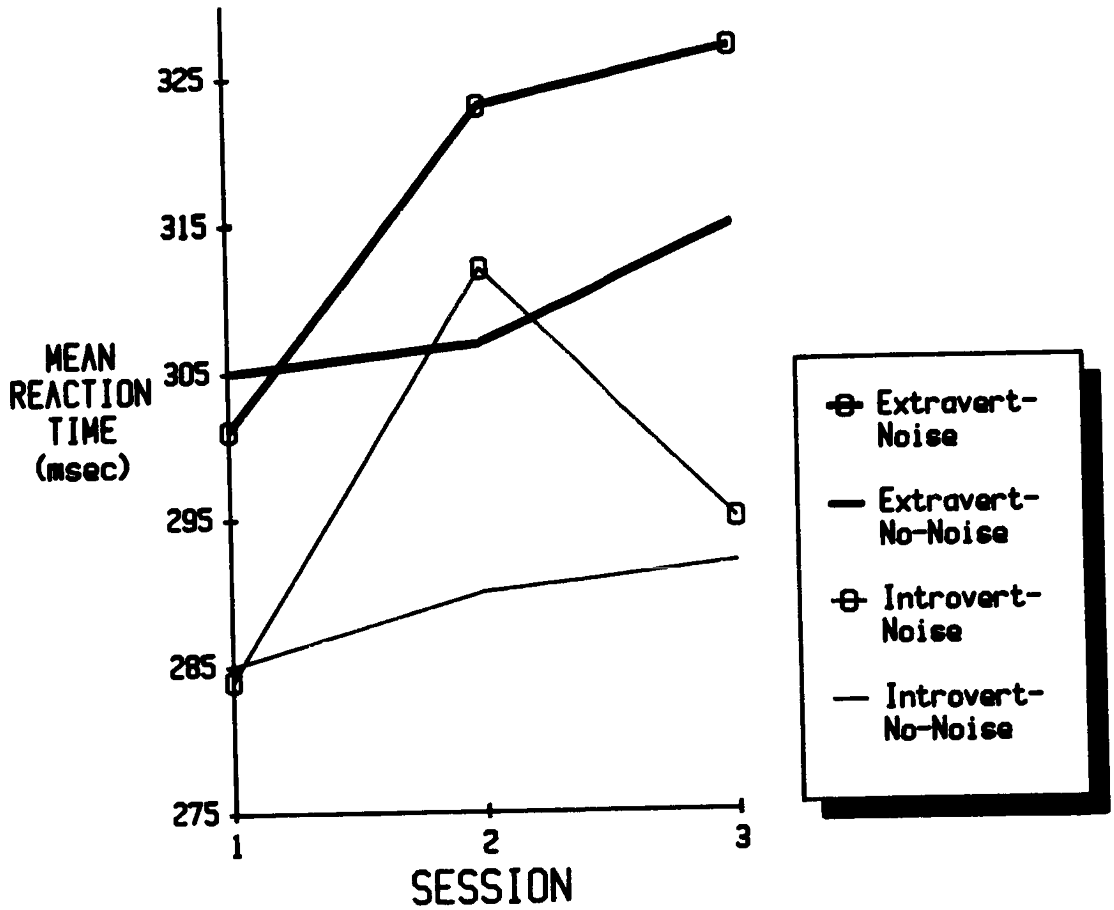


Figure 2: Mean BAER Wave V Latency for Introverts and Extraverts

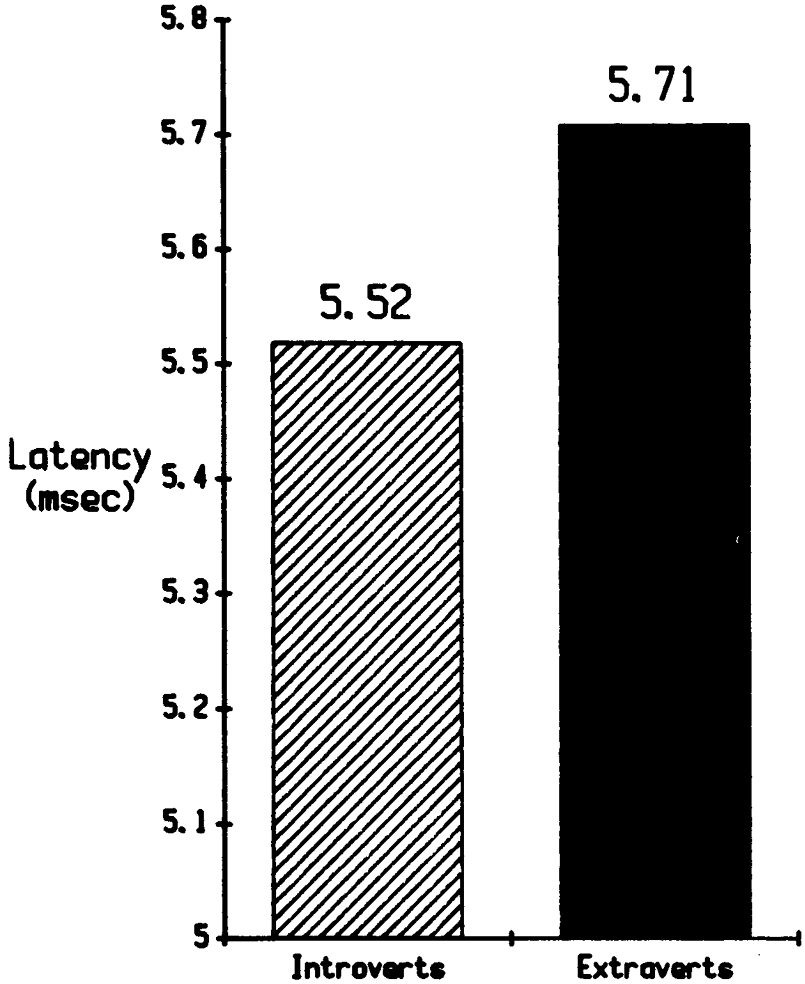
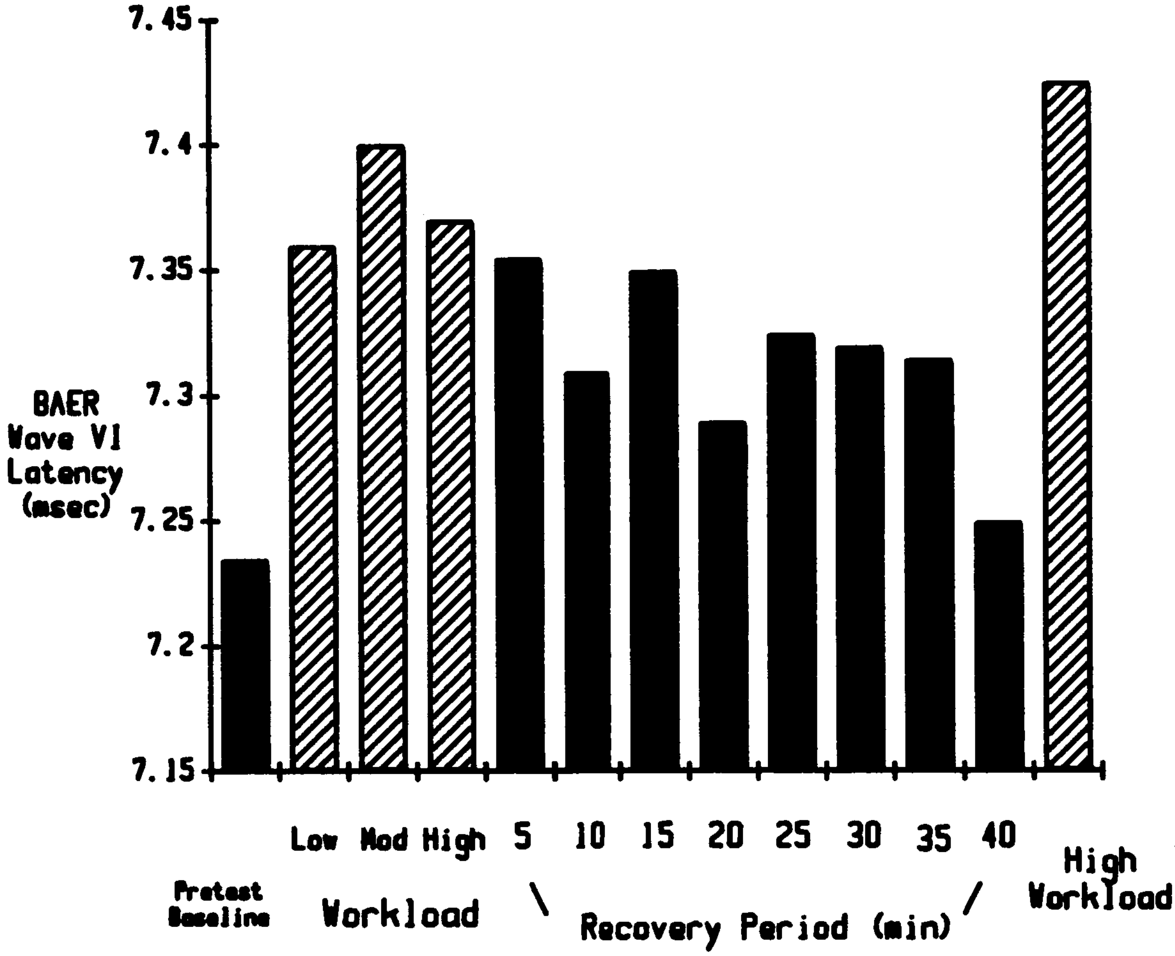


Figure 3: Mean BAER Wave VI Latency for Baseline, Workload, and Recovery Conditions



N88-23395 525-53 140813 14P.

VOICE STRESS ANALYSIS

Malcolm Brenner
National Transportation Safety Board
Washington, D. C.

NN052488 (2)

and

Thomas Shipp
Speech Research Laboratory
Veterans Administration Medical Center
San Francisco, California

VA622374 (1)

A voice measure of the speaker's physiological state has unique applications in the aerospace environment. Unlike other physiological measures, a voice measure is unobtrusive and does not require attaching any equipment to the person being tested. It can be employed in cockpit and spacecraft settings without interfering with ongoing activity and, if used on radio-transmitted speech, might be employed without any additional equipment in the flight environment. A voice measure can also be used on recorded speech as, for example, in accident investigation to determine the relative stress levels of different statements by the flightcrew for information relevant to human performance issues in the investigation. For the purposes of this paper, the term "stress" is used to mean changes in physiological state that result from changes in workload demands.

The aerospace community has been active in research on voice stress analysis (refs. 1 and 2). Although several aspects of the voice have been defined that appear to respond to psychological stress, it remains unclear from the research literature whether such voice changes are sufficiently robust to allow for practical assessment. Practical applications would probably require a single voice measure that is reliable across subjects and situations or, alternately, a battery of voice measures that could be applied to each individual subject and produce a reliable profile of that individual's response to stress.

Research reported in this paper was supported by the School of Aerospace Medicine, Brooks Air Force Base, Texas. It was executed at the Speech Research Laboratory, Veterans Administration Medical Center, San Francisco, California.

VOICE STRESS ANALYSIS

The present paper reports on a research program that is examining issues related to practical voice assessment. The first part of the program was to identify those candidate voice measures from the available research literature that displayed the greatest promise of responding to psychological stress changes. Eight such measures were identified. The second part of the program was to execute an original laboratory experiment that involved clear physiological changes on the part of the subjects within the type of stress range that might be encountered in routine aerospace activity (as opposed to the higher stress range typically encountered in emergency situations from which much of the scientific voice information has been demonstrated). The experiment employed an aviation-like tracking task, varying both task difficulty and monetary incentives.

The third part of the research program was to automate the eight candidate voice measures and compare their responses within the laboratory data to those of traditional physiological measures such as heart rate. This part of the research program is partially complete, with five of the candidate measures automated, and this paper reports the initial results of this effort.

CANDIDATE VOICE MEASURES

Eight candidate voice measures were determined that, it was believed, showed the greatest promise of responding to psychological stress. The choice of these measures was assisted by a comprehensive literature review completed recently for the Naval Air Test Center (ref. 1) and by the authors' familiarity with recent developments in the voice stress area.

The eight candidate measures are

1) Fundamental frequency (pitch). Under stress, there may be an increase in the fundamental frequency of the voice. Fundamental frequency, which may reflect the physical tension of the vocal muscles, is among the most frequently cited voice indices of stress. In emergency situations an increase in fundamental frequency may be universal (refs. 3, 4 and 5).

2) Amplitude (loudness). Under stress, there may be an increase in the amplitude of the voice. This change would probably reflect an increased air flow through the lungs that often occurs under stress.

3) Speech rate. Under stress, there may be an increase in speech rate. This change would be related to a general speeding up of cognitive and motor processes that often appears under stress.

4) Frequency jitter. Under stress, there may be a decrease in jitter of the voice fundamental frequency. Jitter is the minute variability which occurs in the spacing of the fundamental frequency periods (when measured on a cycle-by-cycle basis). It represents a subtle aspect of audible speech that can be difficult to measure precisely (ref. 6). Lieberman (ref. 7) proposed that jitter decreases in response to psychological stress, and there is recent supporting evidence (ref. 3).

5) Amplitude shimmer. Under stress, there may to be a decrease in shimmer of voice amplitude. Shimmer is the cycle-by-cycle variability in the amplitude pattern (and is the equivalent measure to amplitude that jitter is to frequency). Although no literature relates shimmer to psychological stress, it seems reasonable from theoretical considerations that it might follow a pattern similar to that of jitter.

6) PSE scores. Under stress, there may be an increase in scores determined from the Psychological Stress Evaluator (PSE). The PSE is the best-researched of a series of commercial voice devices sold for lie detection. There is substantial evidence that the PSE is not valid for lie detection (refs. 8 and 9), a questionable application for any stress measure that requires subjective determinations by the person administering the test to infer the presence of lying (ref. 10). However, there is also evidence that the PSE-derived scores may respond to simple manipulations of stress (refs. 11 and 12).

7) Energy distribution. Under stress, there may be an increase in the proportion of speech energy between 500 and 1000 Hz. Scherer (refs. 13 and 14) provides evidence for this effect.

8) Derived measure. Under stress, there may be a reliable increase in a derived measure that statistically combines other measures described above. This approach has been advanced by Brenner (ref. 15), who uses the "improper linear model" of Dawes (ref. 16) to provide a simple statistical combination of component speech measures. In theory, the derived measure should then reflect any unusual changes within the same speaker's voice on one or many component measures. In a recent judicial decision, in the legal case of Hoppie/Gillie v. Cessna, such an approach to voice stress analysis was judged to provide admissible evidence (refs. 17 and 18).

VOICE STRESS ANALYSIS

LABORATORY EXPERIMENT

An experiment was designed that, it was hoped, would provide clear physiological differences within the subjects tested. The experiment employed the tracking task of Jex, McDonnell & Phatak (ref. 19), a highly motivating task requiring good reaction time that has been employed extensively in aerospace research. This task can be varied over a wide range of difficulties, and previous literature has suggested physiological changes in response to task loading on measures drawn from heart, respiration, and EMG data (refs. 20 and 21). For the present experiment, monetary incentives were used along with task loading to help guarantee a clear physiological response.

Heart data were obtained from the subjects during the experiment, and excellent voice recordings were obtained of the spoken responses in digital format. Preliminary results available at this time indicate a clear direction for the voice measures that have been tested.

Subjects

Seventeen males, ranging in age from 21 to 35 years old, served as subjects. They were paid $50 plus any monetary incentives won during the experiment.

Procedure

The experiment employed the tracking task of Jex, McDonnell & Phatak (ref. 19) implemented on the Commodore 64 computer. In this task the subject is seated at a CRT display with a manual joystick and attempts to keep a computer-generated triangle at the center of the screen. The triangle moves left and right horizontally in an unpredictable pattern until it touches a left or right boundary on the screen and the trial ends (giving the subject a task similar to balancing a broomstick on a fingertip). A numerical value, the Lambda score, quantifies the mathematical unpredictability of the triangle's gyrations.

Each subject participated at two sessions. At Session 1, the subject was seated in a practice room and trained on the tracking task (25 trials, 10 minute break, 25 trials, ten minute break). At this time subjects performed the "critical" form of the task, in which Lambda was shown on the screen and increased progressively during the trial. The subject attempted to achieve as high a Lambda score as possible before the triangle went out of bounds. To provide speech data, subjects counted aloud on half of the trials. Every ten seconds during the trial, following a computer-generated cueing tone, subjects counted aloud from 90 to 100 as quickly as possible. The counting task was chosen because it causes minimal interference with the tracking task, and the numbers 90 to 100 were chosen because they provide an excellent acoustic pattern with almost continuous voicing.

Following this training, the subject was seated in the laboratory and attached to data recording equipment. Heart rate data were recorded on a multi-channel FM recorder via silver/silver chloride electrode monitors attached to the right and left upper rib areas and base of the neck (the ground electrode). Speech data were recorded via a 1" condensor microphone contained in a custom-modified rubber anaesthesia mask worn by the subject. Speech data were captured digitally in real-time on a laboratory computer at a sampling rate of 10 kHz (the rubber mask also contained a pneumotachograph to measure respiration, and data from this measure are to be described in future papers).

Following a warmup period (ten trials of the "critical" task), subjects performed the "sub-critical" form of the tracking task. In this form the Lambda score, not shown on the screen, was fixed at a specific level of difficulty. The subject's task was to keep the triangle centered for as long as possible up to ninety seconds. On some trials the Lambda score was "easy" (Lambda = 0.9), on some trials "difficult" (Lambda = 90% of the subject's best practice score, median of five trials), and on some trials "moderate" (Lambda = 75% of the subject's best practice score). Each subject performed two trials at each difficulty level. Finally, the subject rested for fifteen minutes, provided baseline measures, and was dismissed. The purpose of Session 1 was training and familiarization, and none of the data collected at Session 1 were analyzed.

At Session 2, several days later, the subject was again seated in the laboratory and attached to data recording equipment. The subject performed a warmup procedure (ten trials of the "critical" task). The subject then performed several trials of the "subcritical" task, both "easy" and "difficult", and these trials represent the principal source of data for the experiment. For these trials, the subject was offered monetary bonuses. On easy trials (Lambda = 0.9) the subject was offered _two dollars_ if he could complete a successful ninety second trial within two attempts. All subjects performed perfectly on the first attempt. On difficult trials (Lambda = 90% of best practice score) the subject was offered _fifty dollars_ if he could complete a successful ninety-second trial within two attempts. Those subjects who failed at this bonus were offered _forty-five dollars_ and two attempts to complete a slightly less difficult task (Lambda = 85% of best score). All subjects succeeded by the end of this second bonus (median Lambda value = 4.2). The order of easy and difficult presentations was counterbalanced across subjects.

To complete Session 2, the subject rested for fifteen minutes and provided baseline measures. The subject was debriefed, paid, and dismissed.

Data Reduction

An automated program was prepared for data reduction related to five of the automated speech measures. The extraction of these parameters was based on algorithms and software developed by E. Thomas Doherty, Ph.D., of the Speech Research Laboratory, Veterans Administration Research Laboratory, San Francisco, California. Dr. Doherty also served as a consultant on this project, and technical details of the analysis program will be provided in other reports.

The automated program inputs recorded speech at slow speed, segmenting it into speech periods and removing the silent periods between syllables and words. The program outputs automated measures for five of the candidate speech measures: fundamental frequency, amplitude, speech rate (ie. total time to speak the ten numbers), jitter, and shimmer.

Results

Data analysis applied to three trials from Session 2 for each subject: the successful "difficult" trial on which the subject won \$50 or \$45; the successful "easy" trial on which the subject won \$2; and a baseline trial on which the subject simply counted. Speech data on each trial consisted of nine repetitions of the numbers 90 to 100.

Figure 1 displays heart rate data. Average heart rate was 83 bpm on the baseline trial, 88 bpm on the easy trial, and 100 bpm on the difficult trial ($F\ (2/32) = 22.1$, $p<.001$). An analysis-of-variance test proved highly significant for the overall difference between difficult and easy ($F\ (1/32) = 21.2$, $p<.001$), and 16 of the 17 subjects showed a higher average heart rate on the difficult treatment than on the easy treatment (sign test: $p<.001$). Based on the heart rate data, then, the experiment produced a clear physiological response against which the voice measures can be compared.

Speech data are summarized in Tables 1 and 2 and in Figures 2, 3, and 4. The analysis-of-variance values reported in Tables 1 and 2 are for differences between the treatment means (a more complete analysis, treatment x time, has not been completed). The second column of Table 2 ("Number of subjects with predicted effect") represents a sign test.

Amplitude displayed a highly significant relation to the task and, as shown in Figure 2, provided a pattern resembling that of heart rate. Average amplitude increased between the easy and difficult treatments by a magnitude of about 0.07 volts, a change that was clearly measurable but that would be virtually impossible to recognize in normal conversation. Fundamental frequency also increased in response to the task, providing a pattern of results less robust than that of amplitude. Average fundamental frequency varied between the easy and difficult treatments by a magnitude of about 2 Hz., a change that is also negligible in normal conversation.

The speech rate measure provided a marginally significant discrimination of the three treatments. Speech rate also showed the highest consistency across subjects of any of the speech measures.

The jitter measure responded in the predicted direction, but to a marginal degree that produced little statistical effect. This measure is of theoretical interest but, pending the results of a complete analysis, does not appear to respond to the type of stress present on this task. Shimmer also responded with marginal effect, but showed a consistency across subjects that suggests a need for further study.

CONCLUSIONS

Previous literature has reported increases in fundamental frequency, amplitude, and speech rate in the voices of speakers involved in extreme levels of stress (refs. 3, 4, and 5)(and these changes are among the major components of screaming). What seems remarkable about the present results is that the same changes appear to occur in a regular fashion within a more subtle level of stress that may be characteristic, for example, of routine flying situations. This evidence adds confidence that these changes reflect some valid underlying physiological response of the human speech system.

The results of our experiment replicate exactly those reported recently by Griffin & Williams (ref. 22). Working in an aircraft simulator setting, they found that increases in speech amplitude, fundamental frequency, and speech rate appeared in the subjects' speech in response to increased workload demands. The combined evidence of the experiments helps establish these three voice measures as parameters for aerospace applications.

In our research, none of the individual speech measures performed as robustly as did heart rate. An area of active future interest is to develop a single derived speech measure, drawing information from several component speech aspects, and to compare the performance of this measure with that of a measure such as heart rate. Another area of future interest is the possibility of developing a convenient and even real-time assessment technique, especially given the current explosion in automated speech processing technology. Voice stress analysis is maturing as a research area, and we urge our colleagues to consider voice response in their thinking about mental-state estimation.

REFERENCES

1. Biers, D. W.: Voice Stress as a Workload Measure: Review of the Literature and Recommendation for Further Study. Final Report, Contract N00421-83-D-0027, Naval Air Test Center, Patuxent River, Maryland, March 1984.

2. Cannings, R.: Speech Patterns and Aircraft Workload. In: Survey of Methods to Assess Workload. NATO AGARDOGRAPH 246, 115-129, August 1979.

3. Brenner, M., Shipp, T., Doherty, E. T., & Morrissey, P.: Voice Measures of Psychological Stress--Laboratory and Field Data. In: I. R. Titze & R. C. Scherer (Eds.) Vocal Fold Physiology, Biomechanics, Acoustics, and Phonatory Control. Denver: The Denver Center for the Performing Arts, 1985.

4. Kuroda, I., Fujiwara, O., Okamura, N., & Utsuki, N.: Method for Determining Pilot Stress Through Analysis of Voice Communication. Aviation, Space, and Environmental Medicine, 47, 1976, 528-533.

5. Williams, C. E., & Stevens, K. N.: Vocal Correlates of Emotional States. In J. K. Darby (Ed.). Speech Evaluation in Psychiatry. New York: Grune and Stratton, 1981.

6. Doherty, E. T., & Shipp, T. Tape Recorder Effects on Jitter and Shimmer Extraction. Unpublished Manuscript, Speech Research Laboratory, Veterans Administration Medical Center, San Francisco, California, 1987.

7. Lieberman, P.: Perturbation in Vocal Pitch. Journal of the Acoustical Society of America, 33, 1961, 597-603.

8. Brenner, M., & Branscomb, H. H.: The Psychological Stress Evaluator; Technical Limitations Affecting Lie Detection. Polygraph, 8, 1979, 127-132.

9. Horvath, F.: Detecting Deception: The Promise and the Reality of Voice Stress Analysis. Journal of Forensic Sciences, 27, 1982, 340-351.

10. Lykken, D. T.: Psychology and the Lie Detector Industry. American Psychologist, 29, 1974, 725-739.

11. Brenner, M., Branscomb, H. H., & Schwartz, G. E.: Psychological Stress Evaluator--Two Tests of a Vocal Measure. Psychophysiology, 16, 1979, 351-357.

12. Schiflett, S. B., & Loikith, G. J.: Voice Stress Analysis as a Measure of Operator Workload. Patuxent River, Maryland: Naval Air Test Center, TM-79-3-SY, March 1980.

13. Scherer, K. R.: Speech and Emotional States. In J. K. Darby (Ed.). Speech Evaluation in Psychiatry. New York: Grune and Stratton, 1981.

14. Scherer, K. R.: Vocal Indicators of Stress. In J. K. Darby (Ed.). Speech Evaluation in Psychiatry. New York: Grune and Stratton, 1981.

15. Brenner, M.: Analyzing Voice Tapes for Psychological Stress. Paper presented at the Annual Southern Methodist University Air Law Symposium, Dallas, Texas, 1986.

16. Dawes, R. M.: The Robust Beauty of Improper Linear Models in Decision Making. American Psychologist, 34, 1979, 571-582.

17. Brenner, M.: Affidavit in Hoppie/Gillie v. Cessna. United States District Court, Missoula, Montana, May 1986.

18. Hollien, H.: Affidavit in Hoppie/Gillie v. Cessna. United States District Court, Missoula, Montana, June 1986.

19. Jex, H. R., McDonnell, J. D., & Phatak, A. V.: A "Critical" Tracking Task for Manual Control Research. IEEE Transactions on Human Factors in Electronics, HFE-7, 1966, 138-145.

20. Derrick, W. L., & Wickens, C. D.: A Multiple Processing Resource Explanation of the Subjective Dimentions of Operator Workload. Technical Report EPL-84-2/ONR-84-1, Engineering-Psychology Research Laboratory, University of Illinois at Urbana-Champaign, February 1984.

21. Jex, H. R., & Allan, R. W.: Research on a New Human Dynamic Response Test Battery. Paper presented at the Sixth Annual Conference on Manual Control, Wright-Patterson AFB, Ohio, April 1970.

22. Griffin, G. R., & Williams, C. E.: The Effects of Different Levels of Task Complexity on Three Vocal Measures. Aviation, Space, and Environmental Medicine, In Press (December, 1987).

Table 1. Differences between the treatment means for the five voice measures (analysis of variance).

	F (2/32)
Fundamental Frequency	7.1**
Amplitude	10.2***
Speech Rate	3.1*
Jitter	0.1
Shimmer	1.3

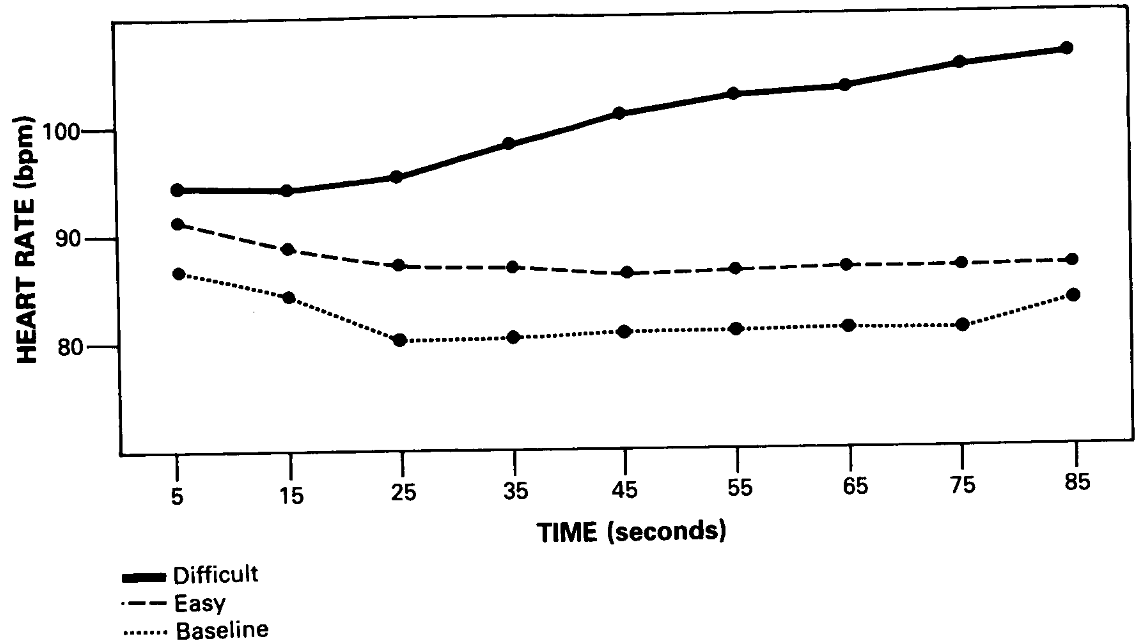
* $p<.10$
** $p<.01$
*** $p<.001$

Table 2. Differences between the easy and difficult treatment means for the five voice measures (analysis of variance/sign test).

	F (1/32)	Number of subjects with predicted effect
Fundamental Frequency	2.9*	10/17
Amplitude	5.0**	13/17**
Speech Rate	2.5	14/17***
Jitter	0.1	9/17
Shimmer	0.7	13/17**

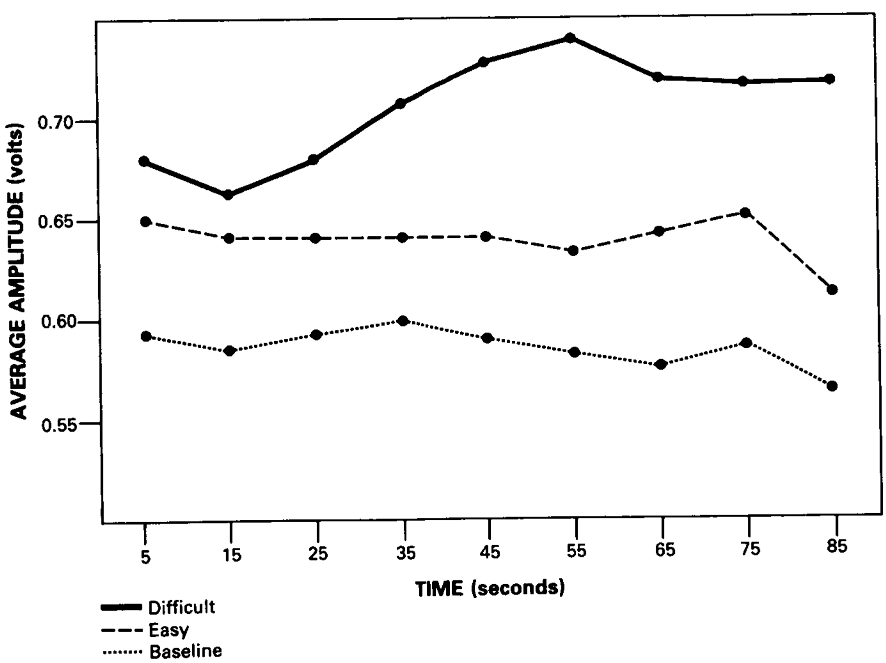
* $p<.10$
** $p<.05$
*** $p<.005$

Figure 1

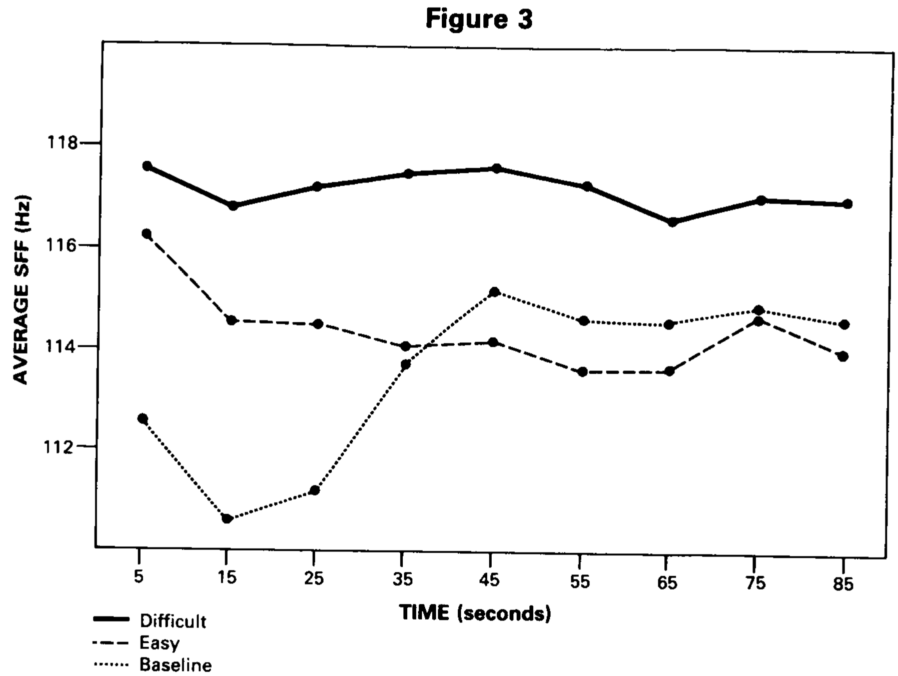


1. Average heart rate over a ninety-second trial as a function of the experimental manipulation.

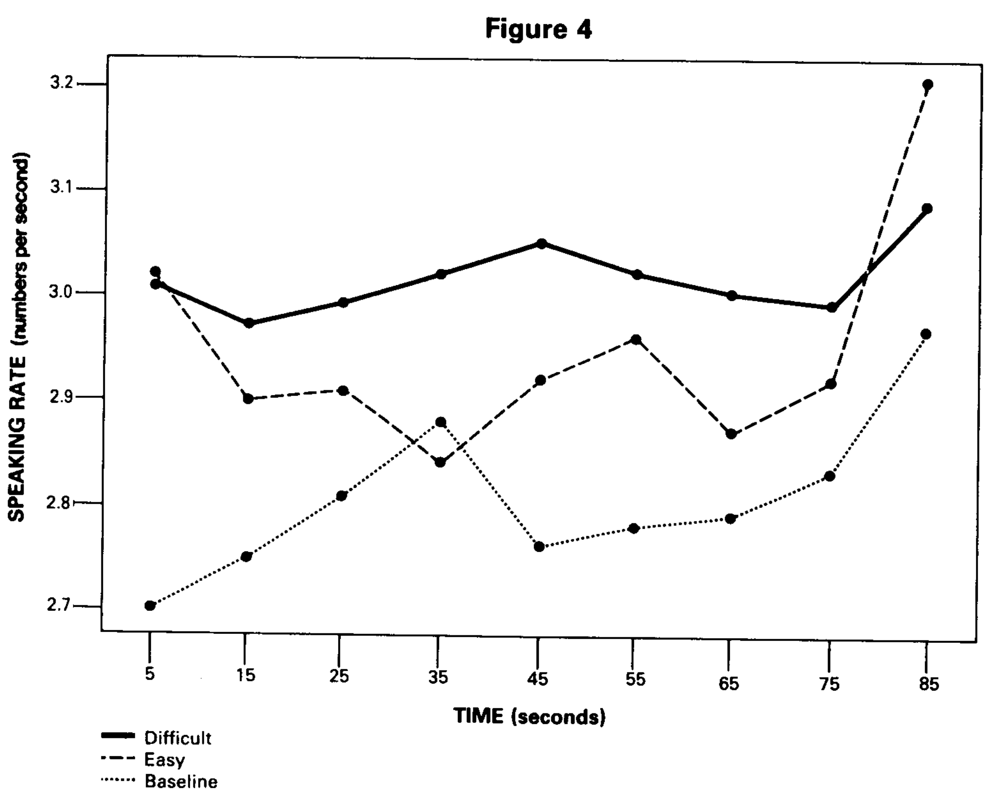
Figure 2



2. Average speech amplitude over a ninety-second trial as a function of the experimental manipulation.



3. Average speech fundamental frequency over a ninety-second trial as a function of the experimental manipulation.



4. Average speech rate (while speaking) over a ninety-second trial as a function of the experimental manipulation. The subject recited the numbers 90-100 every ten seconds.

N88-23396 S26-54
140814
4P
SD 261436

DEVELOPMENT OF A C3 GENERIC WORKSTATION: SYSTEM OVERVIEW

David R. Strome
Systems Research Laboratories, Inc.
Brooks AFB, TX

The mission of the U.S. Air Force School of Aerospace Medicine (USAFSAM) is to support and enhance Air Force Capabilities and Operations through programs across the spectrum of aerospace medicine, education, and research and development. The Crew Performance Laboratory (CPL) of the Aerospace Research Branch, Crew Technology Division is responsible for developing, evaluating, and employing performance measures to allow the assessment of aircrew performance in a variety of environments. The measures include psychophysiological measures, workload assessment tasks, tests of cognitive performance and subjective questionnaires. The environments include chemical defense and performance-altering drugs, sustained operations and stressful situations (altitude, gravitation, hypoxia, disorientation).

One problem that has ramifications in all military services is the effect of selected drugs on human performance at tasks that require decision-making in complex environments and/or under sustained or continuous operations. In each of these situations a decrement in performance when optimal performance is demanded would have disastrous consequences. The CPL has worked closely with the Tri-Service Joint Working Group on Drug Dependent Degradation of Military Performance (JWGD^3 MIL PERF) to develop a facility for evaluating performance in aircrews subjected to chemical defense protection drugs and antihistamines in a complex decision-making command, control and communications (C3) environment. This C3 system is housed in the Aircrew Evaluation Sustained Operations Performance (AESOP) facility which was designed to accommodate sustained operations research.

The following systems, which are based on the proven simulation technology currently in use at the Naval Air Test Center (NATC), Patuxent, MD, will comprise the initial C3 environment and provide flexible, reconfigurable integration (Figures 1-3):

1. A cluster of two VAX 11/780 (Digital Equipment Corporation) computers and peripherals with shared multi-port memory that control scenario presentation and collect performance and physiology data.

2. Four C3 generic workstations configured to realistically simulate, both physically and functionally, the model selected for the simulations.

3. Two VTR-6050 (VOTAN Corporation) speech synthesis/recognition systems under computer control.

4. A state-of-the-art audio distribution communications system, including:

   a. A multi-channel audio recording system.
   b. A white-noise generator.
   c. Simulated radio frequency and intercom channels.

d. An experimenter's control console.

5. All software systems to accomplish the scenario presentations.

The C3 model chosen to investigate this question is the Weapons Director (WD) on an Airborne Warning and Control Systems (AWACS) aircraft. The WD is one of a team of individuals that provides command and control for friendly aircraft in a potentially hostile environment. Specifically the WD locates, identifies and tracks aircraft, controls weapons against targets, ensures expeditious recovery of aircraft, coordinates with internal and external agencies on mission matters and accomplishes tasks assigned by the Senior Director.

Since his function is primarily control the WD has the largest subset of crewstation displays and functions and a considerable communications workload. The simulations presented to subjects with the C3 generic workstations will be both graphics and communications intensive. A state-of-the-art Audio Distribution System Network (ADSN) is currently being developed to provide a realistic simulation of the AWACS communications environment.

Realistic graphic and tabular information will be presented using a CX1500 high resolution graphics subsystem (Chromatics, Inc.) under VAX control. Input to controlling software will be via console switches, trackball and keyboard as in the real-life environment. Switch actions will be recorded with 1 msec resolution. Other performance data will be collected as specified. A physiological data acquisition system with up to 44 channels will be completed and installed into one of the two VAX systems. All data will be time stamped for correlation with scenario events.

The generic workstations, computers, ADSN and speech synthesis units will be combined in a fully integrated network.

All systems must be fully compatible with those that oversee and present the scenarios. Computer control will be affected through the timed event blocks in the scenario data file. This will include transfer of digitized data between the VTR-6050 and the VAX. The difficulty of the operator's task will be changed through modifying the scenario.

In conclusion, the integrated C3 generic workstation facility will provide a powerful, flexible tool for the collection and analysis of data related to aircrew team performance. Complex decision-making environments simulating real situations can be generated for short-term studies or sustained/continuous operations. Performance, physiological and speech data will be collected and analyzed for individuals and teams of individuals.

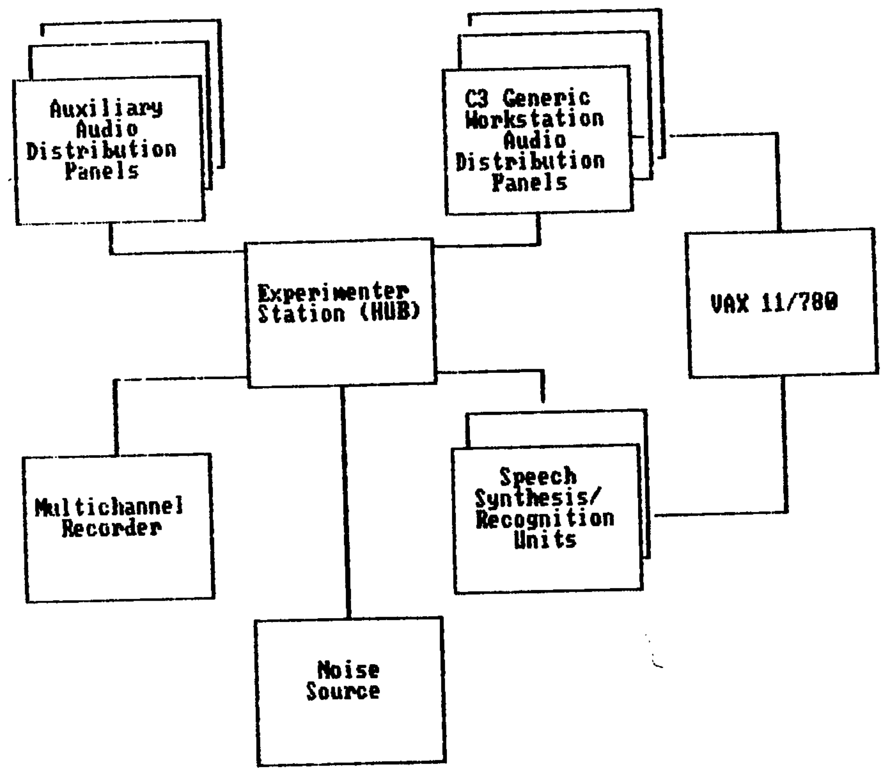


Figure 2. Audio Distribution System Network.

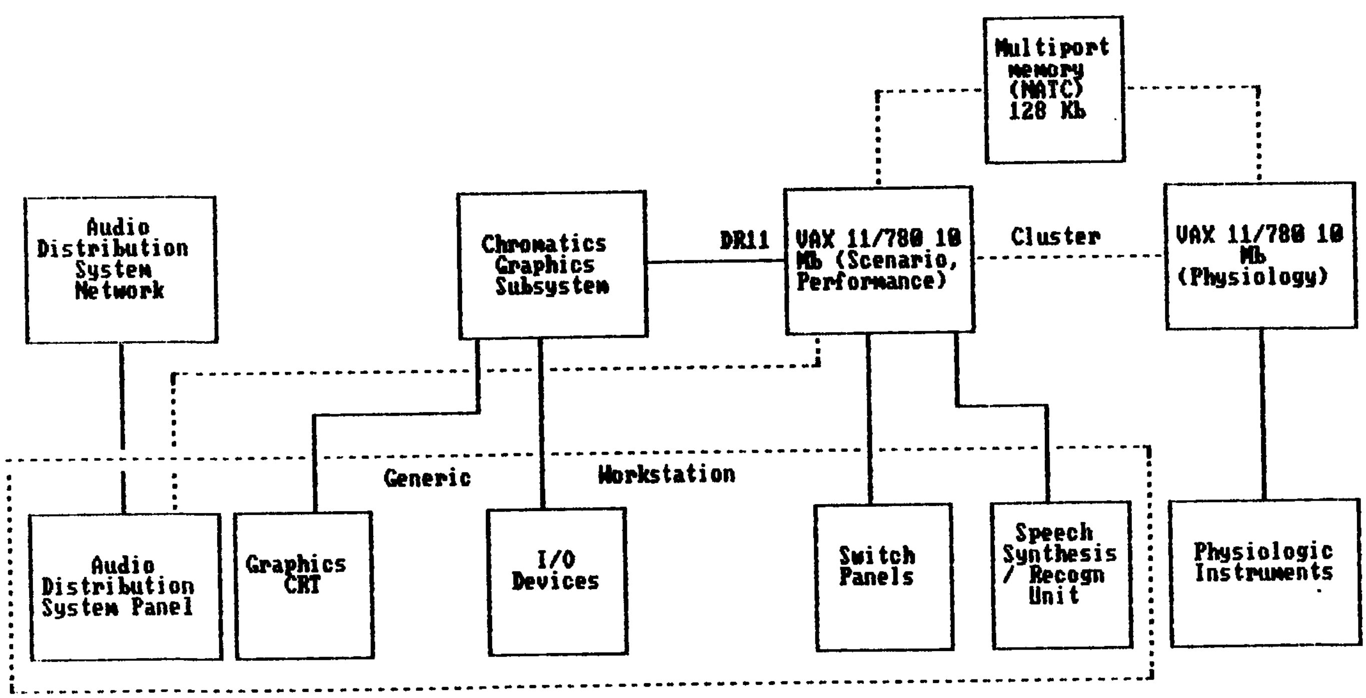


Figure 3. The Integrated C3 Network.

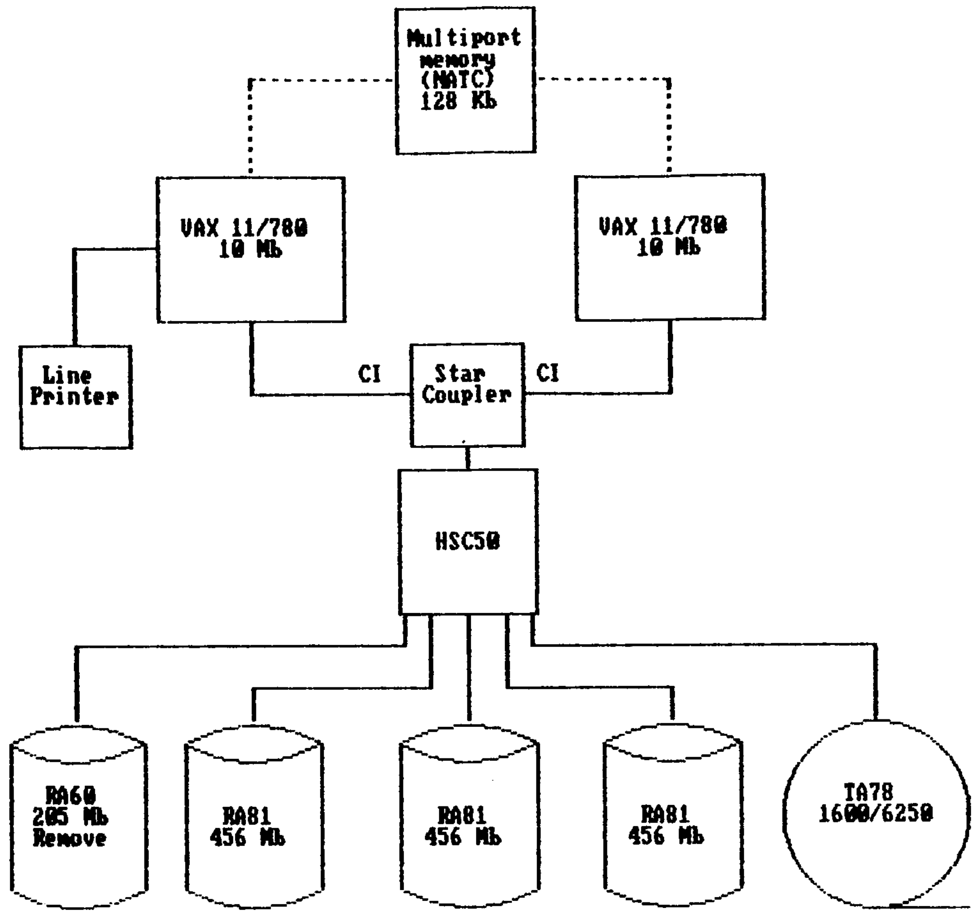


Figure 1. The USAFSAM Crew Performance Laboratory Computer System.

S27-54
N88-23397
140815
4P.

C^3 GENERIC WORKSTATION: PERFORMANCE METRICS AND APPLICATIONS

Douglas R. Eddy
NTI, Incorporated
Brooks AFB, TX

N8389972

ABSTRACT

When a researcher studies complex decision-making tasks in the laboratory, he records a few performance measures and a few physiological measures at most. This presentation describes the large number of integrated dependent measures available on the C^3 generic workstation under development in the Aerospace Research Branch at Brooks AFB. In this system, embedded communications tasks will manipulate workload to assess the effects of performance enhancing drugs (sleep aids and decongestants), work/rest cycles, biocybernetics, and decision support systems on performance. Task performance accuracy and latency will be event coded for correlation with other measures of voice stress and physiological functioning. Sessions will be videotaped to score non-verbal communications. Physiological recordings include spectral analysis of EEG, ECG, vagal tone, and EOG. Subjective measurements include SWAT, fatigue, POMS and specialized self-report scales.

The Aerospace Research Branch will use the system primarily to evaluate the effects on performance of drugs, work/rest cycles, and biocybernetic concepts. We will also develop performance assessment algorithms including those used with small teams. This system provides a tool for integrating and synchronizing behavioral and psychophysiological measures in a complex decision-making environment.

TOPICS

DEPENDENT MEASURES

- BEHAVIORAL
- PHYSIOLOGICAL
- SUBJECTIVE

INDEPENDENT VARIABLES

APPLICATIONS

DEPENDENT MEASURES: BEHAVIORAL

- ACCURACY
  - OMITTED TASKS
  - INCOMPLETE TASKS
  - FALSE ALARMS
  - SEQUENCING ERRORS
  - MISCOMMUNICATIONS
- LATENCY
  - VISUAL
    - TARGETS
    - TEXT
  - AUDITORY
    - COMMUNICATIONS
    - ALARMS
- ACTIVITY
- AUDIO
- VIDEO
- TEAM

DEPENDENT MEASURES: PHYSIOLOGICAL

- ELECTROENCEPHALOGRAM
  - COMPRESSED SPECTRAL ARRAYS
  - STEAD-STATE EVOKED POTENTIAL
  - COHERENCE
  - NEUROELECTRIC PATTERNING (GEVINS)
  - OTHER
- ELECTROCARDIOGRAM
  - VAGAL TONE (PORGES)
- ELECTROOCULOGRAM
  - EYE BLINK (STERN)
  - SACCADES
- VOICE
  - STRESS
  - RECOGNITION
- ACTIVITY
- OTHERS

DEPENDENT MEASURES: SUBJECTIVE

- SWAT
- FATIGUE
- POMS
- OTHERS

INDEPENDENT VARIABLES

WORKLOAD USING EMBEDDED TASKS (SHINGLEDECKER)

DRUGS

- CHEMICAL DEFENSE
- PERFORMANCE ENHANCEMENT
    - SLEEP AIDES
    - DECONGESTANTS

SUSTAINED OPERATIONS

CONTINUOUS OPERATIONS

BIOCYBERNETIC FEEDBACK

DECISION SUPPORT SYSTEMS

OTHER

APPLICATIONS

CHEMICAL DEFENSE DRUG EVALUATIONS

SPACE STATION AND MISSILE SILO WORK/REST CYCLES

BIOCYBERNETIC COMMAND AND CONTROL

MENTAL STATE ASSESSMENT IN COMPLEX DECISION MAKING

- CORRELATION WITH UTC-PAB MEASURES

TEAM PERFORMANCE ASSESSMENT

- COMMUNICATIONS
- COORDINATION
- EFFICIENCY
- REORGANIZATION
- WORKLOAD BALANCING

VOICE RECOGNITION ALGORITHM DEVELOPMENT FOR

- COMPLEX ADAPTIVE TASK MANIPULATIONS
- COMPLEX TASK AUTOMATED TRAINING

OTHER

ATTENDEE LIST

Mr. Theodore B. Aldrich
Anacapa Sciences Corporation
P.O. Box 489
Ft. Rucker, AL 36362
205-598-6326

Dr. Robert A. Alkov
Aeromedical Division,
Naval Safety Center
Naval Air Station
Norfolk, VA 23511
804-444-7341

Dr. Murray Alpert
Department of Psychiatry
NYU Medical Center
550 First Avenue
New York, NY 10016
212-340-5716

Ms Ruth Arnegard
Department of Psychology
Old Dominion University
Norfolk, VA 23529-0267
804-440-4439

Dr. Richard W. Backs
Human Performance Laboratory
E422 107/2/ms10
MDAC STL
P.O. Box 516
St. Louis, MO 63166
314-233-7312

Lt. Commander Guy Banta
Naval Medical Res. & Dev. Command
National Capitol Region
Code 404, Bldg 1, 11th Floor
Bethesda, MD 20814
202-295-1423

Dr. Robert Bateman
Boeing Military Airplane Company
MS K 76-23
Wichita, KS 67277-7730
316-526-2711

Dr. Sarah J. Beaton
Center for Ergonomics
Research and Training
Department of Psychology
Old Dominion University
Norfolk, VA 23529-0267
804-440-4227

Dr. Richard Bloch
Drawer A
Eastern State Hospital
Williamsburg, VA 23187
804-253-5478

Dr. Timothy L. Boaz
Dept. of Clinical Psychology
Box J-165, JHMHC
University of Florida
Gainesville, FL 32610
904-335-8403

Mr. Gregory A. Bonadies
PRC Kentron
303 Butler Farm Road, Suite 100
Hampton, VA 23666
804-865-3348

Dr. Malcolm Brenner
NTSB
TE-50
800 Independence Avenue, S.W.
Washington, D.C. 20594
202-382-6743

Mr. Daniel W. Burdette
PRC Kentron
303 Butler Farm Road, Suite 100
Hampton, VA 23666
804-865-3348

Mr. Evan Byrne
University of Maryland
College Park, MD 20742
301-454-5049

Mr. Joe Cammarota
Code 6023, Naval Air Dev. Center
Warminster, PA 18974

Dr. Patricia A. Casper
Dept. of Psychological Sciences
Purdue University
West Lafayette, IN 47907
317-494-8667

Dr. Ronald A. Cohen
Neurology Department
Univ. of Mass. Medical Center
55 Lake Avenue
Worcester, Massachusetts 01605
617-856-5664

Dr. Glynn D. Coates
Department of Psychology
Old Dominion University
Norfolk, VA 23529-0267
804-440-4439

Dr. J. Raymond Comstock, Jr.
MS-152E
Crew/Vehicle Interface Research Br.
NASA Langley Research Center
Hampton, VA 23665-5225
804-865-3348 FTS: 928-3348

Ms Shari Converse
Department of Psychology
Old Dominion University
Norfolk, VA 23529-0267
804-440-4439

Mr. Jim Deimler
ANALYTICS
2500 Maryland Rd.
Willow Grove, PA 19090
215-657-4100

Dr. Toni Deprospero
Neuro-Psychological Services
Box 286
Weston, WV 26452
304-269-1210 x389

Mr. Douglas R. Eddy
NTI, Inc.
306 Windcrest Avenue
San Antonio, TX 78239
512-532-5723

Ms Linda T. Fatkin
U.S. Army Laboratory Command
Attn: SLCHE-BR (L. Fatkin)
Aberdeen Proving Ground, MD 21005
301-278-5994

Mr. John F. Garren, Jr.
Division Chief
Flight Management Division
MS 153
NASA Langley Research Center
Hampton, VA 23665

Mr. Alan S. Gevins
EEG Systems Laboratory
1855 Folsom Street
San Francisco, CA 94103
415-621-8343

Dr. Kirby Gilliland
Department of Psychology
University of Oklahoma
Norman, Oklahoma 73019
405-325-4511

Dr. Charles Graham
Midwest Research Institute
428 Volker Blvd.
Kansas City, MO 64110
816-753-7600

Ms Paula Guerette
Department of Psychology
Old Dominion University
Norfolk, VA 23529-0267
804-440-4439

Dr. Virginia A. L. Gunther
Dept. of Psychology
S.U.N.Y. - Binghamton
Binghamton, New York 13903
607-777-4862

Dr. Nora Harrington
ANALYTICS
2500 Maryland Rd., Suite 606
Willow Grove, PA 19090
215-659-8808

Dr. Randall L. Harris, Sr.
MS-152E
Crew/Vehicle Interface Research Br.
NASA Langley Research Center
Hampton, VA 23665-5225
804-865-4685 FTS: 928-4685

Dr. Philip D. Harvey
Department of Psychology
State University of New York
at Binghamton
Binghamton, New York 13901
607-777-4991

Mr. Jack J. Hatfield
Branch Head
Crew/Vehicle Interface Research Br.
MS 152E
NASA Langley Research Center
Hampton, VA 23665

Dr. Richard L. Horst
ARD Corporation
5457 Twin Knolls Road, Suite 400
Columbia, MD 21045
301-596-5845

Dr. Andrew Junker
AFAMRL/HEG
Wright-Patterson AFB, OH 45433
513-255-2558

Dr. Barry H. Kantowitz
Battelle
P.O. Box C-5395
4000 N.E. 41st Street
Seattle, WA 98105-5428
206-525-3130

Dr. Raymond H. Kirby
Director, Center for Ergonomics
Research and Training
Department of Psychology
Old Dominion University
Norfolk, VA 23529-0267
804-440-4227

Capt. Richard A. Koffinke, Jr.
Director, U.S. Army Lab. Command
Human Engineering Laboratory
Attn: SLCHE-CC-LHD
Aberdeen Proving Ground, MD 21005
301-278-5952

Dr. Robert Lysaght
ANALYTICS
2500 Maryland Rd.
Willow Grove, PA 19090
215-657-4100 x164

Mr. David L. Mahaffey
ARD Corporation
5457 Twin Knolls Road
Columbia, MD 21045
301-997-5600

Mrs. Mary L. McManus
PRC Kentron
303 Butler Farm Road, Suite 100
Hampton, VA 23666
804-865-3348

Mr. R. Craig Montero
Department of Psychology
Old Dominion University
Norfolk, VA 23529-0267
804-440-4439

Mr. Richard A. Monty
Human Engineering Lab.
P.O. Box 97
Churchville, MD 21028
301-278-5999

Ms Laura E. Moody
Lockheed-Georgia Co.
D72-23/Z419
86 S. Cobb Dr.
Marietta, GA 30063
404-424-5799

Mr. Samuel A. Morello
Assistant Division Chief
Flight Management Division
MS 153
NASA Langley Research Center
Hampton, VA 23665

Dr. Ben B. Morgan, Jr.
Department of Psychology
University of Central Florida
Orlando, FL 32816
305-275-2216

Ms Ilse Munro
U.S. Army Res. Inst.
of Environ. Medicine
SGRD-UE-HP
Natick, MA 01760-5007
617-651-4822

Mr. Robert C. Munson
ARD Corporation
5457 Twin Knolls Road
Columbia, MD 21045
301-596-5845

Lt. Col. Mark Nataupsky, Ph.D.
MS-152E
Crew/Vehicle Interface Research Br.
NASA Langley Research Center
Hampton, VA 23665-5225
804-865-3917

Dr. Lynn C. Oatman
U.S. Army Human Engineering Lab.
Attn: SLCHE-BR (Dr. Oatman)
Aberdeen Proving Ground, MD 21005
301-278-5968

Dr. James P. B. O'Brien
Tidewater Community College
1700 College Crescent
Virginia Beach, VA 23456
804-427-7205

Dr. Brian O'Donnell
Neurology Department
Univ. of Mass. Medical Center
55 Lake Avenue
Worcester, Massachusetts 01605
617-856-4147

Dr. Robert D. O'Donnell
NTI, Inc.
4130 Linden Avenue, Suite 235
Dayton, Ohio 45432
513-254-3171

Dr. David G. Payne
Dept. of Psychology
S.U.N.Y. - Binghamton
Binghamton, New York 13903
607-777-4616

Ms Gwen Pearson
Department of Psychology
Old Dominion University
Norfolk, VA 23529-0267
804-440-4439

Dr. Leslie Peters
U.S. Army Human Engineering Lab.
Attn: SLCHE-BR
Aberdeen Proving Ground, MD 21005
301-278-5968

Dr. Alan T. Pope
MS-152E
Crew/Vehicle Interface Research Br.
NASA Langley Research Center
Hampton, VA 23665-5225
804-865-4685 FTS: 928-4685

Dr. Stephen W. Porges
Department of Psychology
University of Maryland
College Park, MD
301-454-4230

Dr. Stephen M. Rokicki
HQ. U.S.A.F. Oper. Test & Eval. Center
HQ AFOTEC/OAH
Kirtland AFB, N.M. 87117-7001
505-844-9606

Dr. George M. Samaras
GMS Engineering Corp.
8940-D Route 108
Columbia, MDE 21045
301 995-0508

Dr. Sam Schiflett
USAF School of Aerospace Medicine
USAFSAM/VNBC
Brooks AFB, TX 78235
512-536-3464

Dr. Sid J. Schneider
Behavioral Health Systems, Inc.
200 Main Street
Ossining, New York 10562
914-762-6098

Dr. Gary E. Schwartz
Dept. of Psychology
Yale University
Box 11A
New Haven, CT 06520
203-432-4500/4601

Dr. Richard Schweickert
Dept. of Psychological Sciences
Purdue University
West Lafayette, IN 47907
317-494-6930

Dr. Kelly Shaver
Department of Psychology
College of William and Mary
Williamsburg, VA 23185
804-253-4234

Dr. Clark A. Shingledecker
NTI, Inc.
4130 Linden Avenue, Suite 235
Dayton, Ohio 45432
513-254-3171

Dr. John A. Stern
Washington University
Behavioral Science Lab.
1420 Gratton Street
St. Louis, MO 63104
314-621-4211

Mr. Jim Stokes
ANALYTICS
2500 Maryland Rd.
Willow Grove, PA 19090
215-657-4100

Dr. David R. Strome
SRL, Inc.
P.O. Box 35482
Brooks AFB, TX 78235
512-534-1134

Dr. Richard Thackray
Civil Aeromedical Inst., F.A.A.
P.O. Box 25082
Oklahoma City, OK 73125
405-686-4846

Ms Mary S. Thiessen
General Dynamics /
Ft. Worth Division
M.Z. 1766, P.O. Box 748
Fort Worth, TX 76101
817-763-2780

Ms Colleen Thornton
Department of Psychology
Old Dominion University
Norfolk, VA 23529-0267
804-440-4439

Dr. James T. Townsend
Dept. of Psychological Sciences
Purdue University
West Lafayette, IN 47907
317-494-6861

Dr. Larry C. Walrath
McDonnell Douglas Astronautics Co.
E422/107/2/ms10, MDAC STL
P.O. Box 516
St. Louis, MO 63166
314-234-8682

Dr. Ian Wickram
Hampton Roads College of Medicine
P.O. Box 1980
Norfolk, VA
804-446-8981

Mr. Ronald Yates
Research Physiologist
Workload & Engineering Branch
AAMRL/HEG
Wright-Patterson AFB, OH 45433

Dr. Robert L. Yolton
Pacific University
College of Optometry
Forest Grove, OR 97116
503-357-6151 x2272

NASA
National Aeronautics and Space Administration

Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA CP-2504		

4. Title and Subtitle	
Mental-State Estimation - 1987	5. Report Date: May 1988
	6. Performing Organization Code
7. Author(s): J. Raymond Comstock, Jr., Compiler	8. Performing Organization Report No.: L-16420
	10. Work Unit No.: 505-67-11-01
9. Performing Organization Name and Address: NASA Langley Research Center, Hampton, VA 23665-5225	11. Contract or Grant No.
	13. Type of Report and Period Covered: Conference Publication
12. Sponsoring Agency Name and Address: National Aeronautics and Space Administration, Washington, DC 20546-0001 and Old Dominion University, Norfolk, Virginia 23504	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

The Mental-State Estimation Workshop - 1987 was held June 3-4, 1987, in Williamsburg, Virginia. The workshop was sponsored by the Human Engineering Methods Group, Crew/Vehicle Interface Research Branch, Flight Management Division, NASA Langley Research Center, and the Center for Ergonomics Research and Training, Old Dominion University, Norfolk, Virginia. A total of 78 persons attended the workshop, and they represented the government, corporations, and universities. One purpose of the workshop was to examine the status of the idea that cognitive and emotional processes (or mental states) are reflected in discriminable patterns of physiological response. The workshop was designed to explore the potential of a technology based on this concept and the projected flight management applications for such a technology. The papers presented in the conference proceedings reflect the breadth of techniques and areas of study in the examination of mental states and human performance.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Human performance Mental state Operator state Physiological measures Workload	Unclassified - Unlimited Subject Category 53

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of pages	22. Price
Unclassified	Unclassified	399	A17

NASA FORM 1626 OCT 86

NASA-Langley, 1988

For sale by the National Technical Information Service, Springfield, Virginia 22161-2171